# THE FORENSIC IDENTIFICATION OF CCTV IMAGES OF UNFAMILIAR FACES

JEREMY PAUL DAVIS

DEPARTMENT OF PSYCHOLOGY

GOLDSMITHS COLLEGE

UNIVERSITY OF LONDON

Thesis submitted to the University of London in partial fulfilment of the degree of Doctor of Philosophy

January 2007

**Abstract**

Government and private crime prevention initiatives in recent years have resulted in the increasingly widespread establishment of Closed Circuit Television (CCTV) systems. This thesis discusses the history, development, social impact and the efficacy of video surveillance with particular emphasis placed on the admissibility in court of CCTV evidence for identification purposes. Indeed, a verdict may depend on the judgement by members of a jury that the defendant is depicted in video footage.

A series of 8 experiments, mainly employing a single-item identity-verification simultaneous matching design were conducted to evaluate human ability in this context, using both photographs and actors present in person as targets. Across all experiments, some trials were target absent in which a physically matched distracter replaced the target. Specific features were varied such as video quality, the age of participants, the use of disguise and the period of time between image acquisition and identification session. Across all experiments performance was found to be error prone, even if the quality of the images was high and depicted targets in close-up.

Further experiments examined jury decision making when presented with CCTV evidence and also whether extensive examination of images would aid identification performance.

In addition, evidence may be presented in court by facial structure experts in order to verify the identity of an offender caught on CCTV. Some of these methods were discussed and a software package was designed to aid in the identification of facial landmarks in photographs and to provide a database of the physical and angular distance between them for this purpose. A series of analyses were conducted and on the majority of these, the system was found to be more reliable than humans at facial discrimination.

All the results are discussed in a forensic context and the implications for current legal practices are considered.

**Acknowledgements**

I would like to than all of the people that have helped me in writing this thesis. Thanks in particular go to Tim Valentine for his support and guidance. Thanks also to Steve Nicola for his help in recruiting participants and actors and Alan Pickering for his statistical advice.

Special acknowledgements and appreciation must go to Rob Davis for his contribution in designing and writing the software for the *DigitalFace* computer programme.

I would also like to thank all the people who participated in my studies and particularly Mark Russell and Sam Walker for attending identification sessions.

Many thanks go to the Science Museum for allowing me the use of their facilities, especially Sabiha Foster.

I would also like to acknowledge the contribution and help I received from many other academic, secretarial and technical staff members of the psychology department at Goldsmiths College throughout the course of my research.

I would like to thank all of my family and friends for their support in my endeavours, in particular Jo Locker for helping in all departments.

Finally, I would like to thank the Economic and Social Research Council for supporting me financially for the last few years.

## Table of contents



**Chapter One: Thesis introduction**

**List of Figures**

## List of Tables

# Chapter 1: Thesis introduction

## 1. 0. Introduction

Government and private sector crime prevention and public safety initiatives in recent years have meant that Closed Circuit Television (CCTV) systems are becoming increasingly more prevalent. Surveillance cameras are located throughout urban environments including streets, the factory floor, schools, universities, hospitals, sports stadiums, transport systems, retail centres, residential estates and out-of-town commercial sites. In 2000, the Home Secretary, Jack Straw, declared that those working and living in UK cities are filmed by over 300 cameras daily (J. Parker, 2000). There are no official records as to the number of systems in the UK. However, it was estimated in 2004 that annual expenditure was more than £1 billion, with approximately 4,285,000 cameras sited across the country (McCahill & Norris, 2003a; Norris, McCahill & Wood, 2004). These figures are expected to rise. In his New Year speech in December 2004, the Prime Minister, Tony Blair announced, "the biggest ever expansion of CCTV underway to ensure we spot, catch and convict the criminals" (Number 10, 31 December, 2004). Indeed, in 2005 it was reported in *The Times* that there may already be more than 7,000,000 cameras in place (Irving, 2005).

The UK is believed to have the highest system density in the world (McCahill & Norris, 2003a). However, similar large-scale implementation appears inevitable in other countries (Norris et al., 2004; Sutton & Wilson, 2004). Norris et al. (2004) note that whereas industry analysts were originally anticipating annual sales of approximately 2 million cameras in the USA, costing $1.6 billion by 2001, this had actually expanded to $5.7 billion following the September 11$^{th}$ 2001 attacks. Indeed, it was estimated by the *Washington Post* that there may already be as many as 26,000,000 cameras within the USA (*Washington Post*, 8 October 2005). In this context, the research findings from pioneering studies investigating the effectiveness of CCTV, primarily conducted in the UK have implications for policy development elsewhere. Indeed, the FBI (USA) best practice guidelines for the implementation of CCTV systems and forensic image analyses were all produced in the UK (Scientific Working Group on Imaging Technology (SWGIT), 2005).

The prevalence of CCTV surveillance has given rise to fundamental questions in relation to its efficacy as a crime prevention tool but also its reliability when images are obtained and presented as evidence in court. It is likely to be used more frequently within the criminal justice system, in particular for the identification of those involved in illegal acts. However, research within the field of psychology has highlighted the difficulties involved in the successful identification of unfamiliar people depicted in even in the highest quality images (e.g., Bruce, Henderson, Greenwood, et al., 1999). Parallel sociological studies have demonstrated that CCTV may not necessarily reduce local crime rates (e.g., Gill, Allen, Bryan, et al., 2005).

This thesis discusses the prevalence and expansion of CCTV systems in the UK and the rest of the world. It also examines the sociological impact of large scale surveillance, both in terms of its effect on crime, its acceptance by different groups and ethical issues that have been identified, such as a perceived loss of individual privacy. Technological innovations are also assessed, including the potential integration of high resolution digital networked systems and the development of face and behavioural recognition algorithms. However, the primary topic of investigation is to evaluate how identification evidence from surveillance cameras is used within the criminal justice system in the UK. As such, a series of experiments are reported that were designed to simulate aspects of publicised court cases. The results of these studies have legal implications, especially in cases in which the identity of a defendant is disputed.

### 1. 1. The technological specifications and social implications of CCTV

Implemented and marketed primarily as a crime prevention measure, many UK CCTV schemes have been financed by the Government. Over three-quarters of the Home Office annual crime reduction budget between 1993 and 1996 was dedicated to the installation of surveillance systems (Peace, 1997) and the overall cost to the taxpayer had exceeded £1 billion by 2002 (Farrington & Painter, 2003). Considering this substantial financial outlay, concerns have been raised that the money spent on research into the effectiveness of CCTV has been disproportionably low (Farrington & Painter, 2003). Studies claiming substantial successes have been described as "post hoc shoestring efforts by the untrained and self interested practitioner" (Pawson & Tilley, 1994). Independent studies have tended to find only minor reductions in crime statistics, and in some cases, relative increases compared to control areas (e.g., Welsh & Farrington, 2002, Gill et al., 2005). This may be due to a rise in reported crime, but also to a false sense of security by victims leading them adopt a more vulnerable behaviour. Furthermore, some criminal activity appears to be displaced to neighbouring localities (Flight, van Heerwaarden & van Soomeren, 2003).

Although some of this research has been criticised, it has generated questions as to whether other crime reduction initiatives would be more cost-effective; for instance, an increase in police patrols, prisons or community regeneration programmes (Farrington & Painter, 2003). Nevertheless, CCTV is viewed extremely positively by the public (Gill, Smith, Spriggs, et al., 2003), the police (Brandon, 2003), politicians (Norris & Armstrong, 1998) and businesses (Skinns, 1998); not only because of a belief in it's long term positive impact on crime, but also for making the public feel safer, and in the detection and identification of criminals. Furthermore, Reeve (1998) argues that without a CCTV system, a town centre can be perceived as second-rate, as surveillance acts to support business and leisure activities by the maintenance of a pleasant environment.

The high financial costs associated with CCTV must also be qualified, as the technical specifications of systems vary extensively. Indeed, a proportion of installations are fakes, designed to act as a visual deterrent (McCahill & Norris, 2003a). At the most primitive level a single fixed camera may be directed at a specific area; for instance, a till in a shop. No recording is undertaken and monitoring is rare. Most UK town centre and open-street systems currently use multiple analogue cameras connected to a central control room with a number of specifications to improve picture quality. These include zoom, tilt, pan, night vision, motion detection and infra-red facilities. However, they suffer from being adversely susceptible to changes in environmental conditions, of low image resolution, and are often set well above ground level, resulting in unclear images. Analogue systems also require extensive tape storage facilities and substantial operational manpower. More recently, technologically sophisticated high resolution digital systems have become commercially viable, the management of which is cost effective in comparison to analogue-based CCTV (Bull, 2003). Many cameras can be integrated into a single network and the necessity for physically extensive data storage is reduced, allowing efficient coordinated post-event analysis. Gill and Loveday (2003a) predict that these improvements will act as a more effective deterrent so that future crime evaluation studies produce more positive findings

Graham (1998) also predicted that in the future public CCTV systems will be combined into a single integrated network or "fifth utility" (alongside gas, water, electricity and telecommunications). Systems are also being designed to analyse movement, alert operators to the presence of known criminals and to suspicious behaviour patterns (Webster, 2004). The development of the computer algorithms necessary for these tasks is still in its infancy. However, commentators suggest that when perfected the result will be a "Maximum Surveillance Society" (Norris & Armstrong, 1999, p. 12). The sociological implications of these systems together with the history of CCTV, its placement, the implications on privacy and civil rights and its effectiveness in terms of crime reduction are discussed in Chapter 2.

### 1. 2. Identification evidence and the criminal justice system

Prior to the comprehensive coverage of surveillance cameras, the police when investigating many crimes could only rely on eyewitness identification. However, many studies have shown that the identification of unfamiliar individuals based on memory is fallible (e.g., G.M. Davies, 1996; Wells, 1993), with confidence in false identifications often being quite high (Luus & Wells, 1991; Sporer, Penrod, Read & Cutler, 1995). Eyewitness errors have been identified as one of the primary causes of miscarriages of justice. For instance, Rattner (1988) found that of 205 cases of wrongful conviction in the USA, over 50% were because of mistaken eyewitness identifications. Similar figures were found by Scheck, Neufeld and Dwyer (2000) examining the case histories of 62 previously-convicted but innocent prisoners, exonerated by the ability to present DNA evidence on appeal.

Identification from lineups is the primary evidence in at least 80,000 cases per annum in both the USA and the UK (P. Burton, 2006; Goldstein, Chance & Schneller, 1989). Therefore the specific cases identified above by Scheck et al. (2000) and Rattner (1988) may be isolated examples and not representative of a greater system malaise. However, approximately 20% of identifications from lineups in England and Wales result in the selection of innocent distracters (Valentine, Pickering & Darling, 2003; Wright & McDaid, 1996). Wells, Malpass, Lindsay et al. (2000) do note that until cases of wrongful imprisonment were publicised, the media and the legal system in the USA largely ignored the results from psychological literature concerning eyewitness fallibility. In the UK, legal and media interest into research also occurred following the publication of the Devlin report reviewing 36 misidentification cases (Devlin, 1976). Its main proposal, not accepted by the Government, was that except in extremely rare instances, convictions based on eyewitness evidence alone should cease. However, in the USA and UK, changes to recommended practices have occurred (e.g., UK: Police and Criminal Evidence Act, 1984; USA: Technical Working Group for Eyewitness Evidence, 1999). Following the 'Turnbull Guidelines', judges in the UK are expected to warn juries as to the potential unreliability of eyewitness identification evidence especially if viewing conditions are allegedly limited or poor (*R v. Turnbull and others*, 1976). In these circumstances, if cases are unsubstantiated by other evidence a judge should direct the jury to acquit, a procedure adapted by other common law jurisdictions such as Canada and Australia (Bromby, 2004).

These legal concerns have meant that if CCTV footage is available, greater evidential weight is placed upon it, as there is often no necessity to employ the memory of witnesses. On viewing an offence, CCTV operators can track a culprit until the police arrive to ensure that the correct offender is apprehended. In cases in which the identity of a perpetuator is not contested, video footage can be presented in court for incident verification. However, if the culprit is not immediately apprehended, recordings are also used for identification purposes. In these circumstances, facial images can be matched to the suspect in custody or to a photographic database of known faces. If there is no candidate, some police forces issue CCTV stills as part of a daily online briefing to local officers and in internal journals such as the *Police Gazette*. CCTV images are sometimes made available to the local media, and when crimes are particularly serious can be publicised nationally or even internationally. In each case, the aim is that someone familiar with the perpetrator will make a positive identification.

Software systems have been designed to specifically match individual faces seen on video with databases of faces. However, at present, performance is only better than normal human ability under optimal conditions. When views are incongruent, or images are filmed using different lighting or other environmental conditions, accuracy is far worse (A.M. Burton, Miller, Bruce et al., 2001; P.J. Phillips, Grother, Micheals et al., 2003). Considerable investment is being undertaken to improve these systems, but until empirically substantiated as consistently more reliable than human ability, human observers will still be required to make the final match between a CCTV image and a potential suspect for legal purposes.

**1. 3. The use of CCTV images in court**

Photographic identification evidence has been admissible in the UK since 1864 (*R v Tolson,* 1864; cited in Murphy, 1999). CCTV footage itself was first used in court to provide information about theft from a retail store (*R v Fowden and White*, 1982) and is now regularly used to support other evidence (e.g., *R v Christou,* 1992). If the images and events shown in the footage are unclear it has been used to add weight to eyewitness testimony as it can corroborate statements (e.g., *R v Pattinson and Exley*, 1996). Video footage has also been presented in a courtroom, with jury members encouraged to provide a verdict based on their perception as to whether a perpetrator shown on video is the suspect in the dock. The legal basis for this was tested when an appeal against conviction was submitted (*R v Dodson and Williams,* 1984). The original trial jury had been shown CCTV stills from a bank raid and invited to compare them with the two defendants, which the appeal defense counsel argued amounted to 'dock' identifications. These have been deemed to be undesirable, in eyewitness cases due to potential bias in comparison to pre-trial standardised lineups (*North Yorkshire Trading Standards Department v Williams,* 1994). Moreover, the prosecution counsel had stated that the stills 'clearly revealed' the defendants, inviting conclusions which the defence argued could prejudice jurors' opinions. No corroborating identification evidence was submitted although the court was presented with photographs of one of the accused taken the day after the offence to compare with the CCTV stills.

The appeal was dismissed on the basis that it had been correct for the jury to view the images and that their task required no special expert training. The original judge had also cautioned the jury that photographs do not always provide a good resemblance. The Appeal Court concluded that:

> "so long as the jury - are firmly directed that to convict they must be sure that the man in the dock is the man in the photograph, we envisage no injustice arising from this manner of evaluating evidence with the aid of what the jurors' eyes tell them is a fact which they are sure exists"

Later trials have confirmed the acceptability of juries making decisions on this basis (e.g., *R v Blenkinsop*, 1995) and in one a jury asked for the defendant to stand and turn around to compare his appearance with video footage (*R v McNamara*, 1996).

Opinion and expert evidence as to identity from CCTV has also been permitted. For instance, witnesses previously familiar with a suspect have given evidence after viewing images and making a positive identification (e.g., *R v Grimer*, 1982; *R v Caldwell and Dixon*, 1993), even when the original recording was destroyed (*Taylor v The Chief Constable of Cheshire*, 1987). In these circumstances, identifications are presumed to have the same status as those from eyewitnesses actually present at the incident. The jury can decide how much weight this evidence should be given as witnesses can be cross-examined. Indeed, in *R v Caldwell and Dixon* (1993), three police officers who had initially recognised the suspects from video footage, later selected the same suspects from lineups, adding credibility to their testimony. Nevertheless, there have been eyewitness cases in which close friends or relatives have mistakenly identified an innocent familiar person (e.g., *R v Bowden*, 1993; *R v Thomas*, 1994). Therefore, the Turnbull guidelines are normally applied when evidence of this type is presented (e.g., *R v Campbell*, 1996).

Evidence may also be admissible if an individual claims to have gained specific identification expertise from closely inspecting video footage even if previously unfamiliar with those depicted. In *R v Clare and Peach* (1995) a police officer viewed black-and-white CCTV footage of a football crowd riot more than 40 times, examining stills and evaluating details in slow motion. He also compared this footage with separate colour photographs showing undisputed images of the defendants taken the same day. His testimony was available for cross-examination and the court ruled that due to the time spent scrutinizing the images he had gained a "special knowledge that the court did not possess" and as such had developed an 'ad-hoc' expertise.

Finally, practitioners from different disciplines, including medicine, computer science and art may be invited to present evidence based on their professional expertise. In these circumstances, judges have to decide on the scientific validity of the technique and the authority and experience of the witness as well as to determine whether their presence is necessary. Indeed, experts should not be called if a judge and jury are able to form their own opinion without them (CPS, 2005).

Different methodologies have been employed by experts in different cases. Some have involved the application of *facial mapping* techniques, which entail the measurement of face structure (e.g., *R v Clarke*, 1995; *R v Stockwell*, 1993). One such method involves locating various facial landmarks from which distances or angles are calculated. A comparison can then be made between a CCTV image and a photograph of the defendant to see if these dimensions match. Some research using this type of methodology has been published (e.g., AM Burton, Bruce & Dench, 1993; Mardia, Coombs, Kirkbride, Linney, & Bowie, 1996). However, there does not appear to have been a comprehensive investigation of the distribution of measurements in the population and problems can be encountered if the referent images are not aligned or facial expressions are altered. Although it would normally be accompanied by other supportive evidence this type of testimony has been deemed admissible without further substantiation of identity (e.g., *R v Hookway*, 1999). As such, juries would be directed to draw their own inferences as to the credence of the expert and the evidence.

The legislation concerning the use of CCTV evidence for identification purposes in court in the UK was summarised in a recent reference to the Attorney General by Appeal Court judges (Attorney General's Reference, No. 2 of 2002, 2003). Four scenarios were recommended in which CCTV evidence would be appropriate to assist in establishing the guilt of the accused.

1. If identifications have been made after viewing a video by individuals' previously familiar with a defendant, they may give evidence as a witness for the case even if the footage is unavailable.

2. "Where the photographic image is sufficiently clear, the jury can compare it with the defendant sitting in the dock" (p. 5).

3. A witness not previously familiar with the defendant may spend "substantial time viewing and analysing photographic images from the scene" (p. 6), thus familiarizing themselves with the accused and gaining a special knowledge not possessed by the jury. Identifications by this witness can then be based on the perceived resemblance between these images and an undisputed contemporary photograph of the defendant, which should be made available to the jury.

4. Qualified experts in facial mapping or face structure may provide opinion evidence as to whether the individual captured on video footage is the same as that in a contemporary photograph of the defendant. Again all images should be available to the jury.

However, a later ruling laid out extra conditions which should ideally be met when evidence is provided by facial mapping experts (*R v Gray*, 2003). These included the creation of a national database of facial measurements, similar to that for fingerprints in order that the probability of the occurrence of specific facial features or a combination of those features can be objectively established. The judges did not suggest that evidence from facial mapping experts should be inadmissible. However, without this safeguard they argued that opinions were potentially subjective in nature.

Indeed, in a later Appeal Court ruling, evidence from an expert witness was allowed after using specific equipment that allowed him to 'subjectively' state with 'high probability' that the defendant was depicted in CCTV footage, without giving any indication of the likelihood of occurrence of the specific facial features within the population (*R v Gardner*, 2004). As such, by using his equipment and by the frame-by-frame inspection of the images, he was deemed to be able to provide the jury with opinion evidence of the identity of the person depicted, in the same manner as in *R v Clare and Peach* (1995).

The legal principles on identification evidence from CCTV in the UK would not necessarily apply in different countries. However, newspaper reports indicate that in the USA images are also regularly shown to juries to compare with the defendant as 'proof' of guilt (Grimm, 2006; Treleven, 2006). In Australia, evidence from police officers who claim to be able to recognise an offender would not normally be admissible, as juries are believed to be as capable of making their own decisions as to identity from viewing footage and comparing it to the defendant (*Smith v The Queen, 2001*). However, expert witnesses in facial mapping have been called in some cases (Michaelmore, 2005). In addition, in Canada, evidence from police officers familairising themselves with an individual in CCTV as was described in *R v Clare and Peach* (1995) would not be admissible. However, evidence can be shown to a jury for them to decide on identity (Leaney & Rawlinson, 1988, cited in Mead, 2003).

Even though they may be warned in advance of its potential weaknesses, juries and law officers in the USA have been found to place a particularly high, potentially erroneous credence on eyewitness evidence (Brigham & Bothwell, 1983; Wise & Safer, 2004), especially if the witness is confident (Brigham & Wolfskeil, 1983; Cutler, Penrod & Dexter, 1990). Indeed, in the USA, confidence in eyewitness testimony is regarded as a criterion of accuracy (*Neil v Biggers*, 1972). It is therefore possible that individual jurors might place even greater weight on CCTV evidence, especially if they personally believe that a video image appears to match a defendant in court, or a photograph of the accused taken at about the same time. As such, they would be able to 'see for themselves' the resemblance. It is also possible that regardless of other evidence, verdicts may be rendered on this basis.

Real jury deliberations are conducted in private and information as to how a jury has come to a particular decision is confidential. However, research has attempted to simulate the decision-making processes of juries (e.g., Bornstein, 1999). Jury decision making has been found to conform to Social Decision Scheme (SDS) models of group decision making (e.g., Davis, Kerr, Atkin et al., 1975; Kerr, MacCoun & Kramer, 1996). These propose that final verdicts can be predicted by initial voting patterns, suggesting that despite any minority reservations, if most jury members believe that the accused is shown in CCTV footage it is probable that a guilty verdict will ensue. This scenario is likely to become more common with the expanding prevalence of higher-quality images obtained from digital systems and is one of the primary topics of this thesis.

## 1. 4. The identification of familiar people in CCTV images

The first recommendation as to the admissibility of CCTV evidence in the Attorney General's reference discussed identifications made by individuals familiar with a suspect (Attorney General's Reference, No. 2 of 2002, 2003). The success of media appeals are based on this premise and images of offenders are often shown with the aim of soliciting a positive identification. For example, the Brixton Nail Bomber, David Copeland was identified by a work colleague after images were broadcast on television (Hopkins & Hall, 2000). Two recent studies have demonstrated that the recognition of familiar faces in CCTV images is robust, even if image quality is poor (Bruce, Henderson, Newman & Burton, 2001; Burton, Wilson, Cowan & Bruce, 1999). For instance, Burton et al. (1999) found that university students were 90% correct when recognising lecturers from their own department in poor-quality video. A similar high level of accuracy was found by Bruce et al (2001) using a task in which participants were presented with a series of pairs of facial images. One image in each pair was either a still or moving footage from a poor-quality CCTV system, the other a facial photograph. When participants were familiar with targets, accuracy at identifying those shown in the pair as the same person, or as two different people was extremely high.

However, one potential confounding variable was that the images were shown in context-rich settings, which has been shown to aid recognition (e.g., Young, Hay & Ellis, 1985). In both studies, footage of psychology lecturers in department corridors was presented to participants who were all their students. It is less clear whether accuracy would be so high in an unexpected context. Indeed, isolated cases have been publicised involving errors of familiar-person identification. For instance, close family members of a missing person all wrongly identified a man as their relative filmed by high-quality airport CCTV footage (BBC News, 16 August 2003). There may be confounding explanations for these errors as the initial false identification would have been by someone unknown to the target. Contextual details such as similar clothing may also have contributed. Nevertheless, this case does illustrate that recognition of even highly familiar people is not infallible.

**1. 5. The identification of unfamiliar people in CCTV images**

In contrast to the high recognition rates of familiar faces, identification of unfamiliar people on video has been found to be surprisingly unreliable even when there are no memory demands and the quality of the image is extremely good (e.g., Bruce et al., 1999; Bruce et al., 2001; Henderson, Bruce & Burton, 2001). This finding is of consequence as most incidents caught on CCTV are likely to involve people not known either to operators or to the police. Furthermore, in a courtroom, members of a jury could base their judgements on whether the defendant resembles the offender shown on CCTV footage. As they would be previously unfamiliar with the suspect, this could have serious implications. And yet, this scenario, without the need for further identification evidence forms the second recommendation in the Attorney General's reference (Attorney General's Reference, No. 2 of 2002, 2003).

Two of the studies reported above, directly compared familiar and unfamiliar face matching using poor-quality CCTV images. Burton et al. (Experiment 1) initially exposed participants to a series of stills, finding that familiarity was associated with fewer false alarms and more hits in a later recognition task. Similarly, Bruce et al. (Experiment 1) used a single-item identification-verification matching design with high-quality photographs and low-quality CCTV stills. They found that if participants were familiar with the targets the hit rate was approximately 93%. However, when targets were unfamiliar, performance was reduced to 76%. If targets were presented simultaneously with a distracter, correct rejections remained high but only if the target was familiar (91%). When both were unfamiliar, the correct rejection rate was approximately 55%. These difficulties may partly be due to the poor video quality. However, studies using extremely high-quality images have also been discouraging.

For instance, Bruce et al. (1999; Experiment 1) reported error rates of 30% in a task requiring the matching of male frontal facial high-quality video stills with simultaneously presented frontal high-quality facial photographs among an array of nine distracters. Equivalent false negative error rates were found in target absent trials. When facial expressions or pose differed, accuracy was further reduced. And yet, photographs and films were taken on the same day and the appearance of the actors would not have substantially changed. Disturbingly, the reported values reflected average performance. In one specific trial, 80% of participants were unable to correctly select a target individual from the array. These findings demonstrate that apparently small differences in even high-quality image formats are responsible for a large detriment in performance. Images from CCTV systems are rarely of this quality and similar array studies utilising standard images have found considerably higher error rates (Henderson et al, 2001; Experiments 1 and 2), especially if captured by cameras fixed to high-level pylons (Davies & Thasen, 2000; Experiment 1).

Two different experiments by Henderson et al. (2001) illustrate that even with reduced task demands, face matching performance is still error prone. The first (Experiment 4) utilised a two-alternative forced-choice design in which participants were asked to identify which of two photographs depicted a target actor shown in a video still. One was a picture of the target, one a distracter of similar appearance. Overall 76% of decisions were correct. However, in one trial, approximately one-third of people thought that a still of one actor was more similar to a photograph of a distracter than the actor's own photograph. Confidence in these decisions was consistently high, even when incorrect. In a follow-up experiment (Experiment 5), using a single-item identity-verification design, approximately 45% of participants believed that two images of the same person were of different people. Moreover, 27.5% incorrectly matched the images of two different actors. What is most concerning is that image quality was extremely high. Close-up facial stills were from high-quality television broadcast film and the photographs were professional studio portraits. Forensically, of most concern is the high percentage of false positive results, as these represent scenarios whereby an innocent suspect could be wrongly mistaken for the offender caught on CCTV footage.

An important feature of both of the above studies was the use of relatively small databases of individuals. Bruce et al. (1999) included 160 faces from 200 trainee police officers, whilst Henderson et al. (2001) "searched through several hundred actor-agency photographs" (p. 463), to select appropriate photographs to use as distracters. It therefore appears comparatively easy to construct experimental designs in which errors in identification matching occur. With a larger database, overall error rates may have risen as it should be easier to acquire more distracters resembling the targets. This also suggests that there may be many people in the population who could easily be mistaken for one another.

Further problems in identification may be encountered by the typical positioning of CCTV cameras, often sited above head height with a large field of view, lessening the likelihood of close-up facial images (Davies & Thasen, 2000; Experiment 1). Using this type of image the authors found extremely poor matching performance and suggested this was primarily due to the differences in camera angle between the video footage and photographic target. Indeed, the importance of specificity of viewpoint, expression, and of environmental lighting effects reflect similar results found in face matching (e.g. Bruce et al., 1999; Bruce, Valentine, & Baddeley, 1987; Hill & Bruce, 1996) and recognition studies (e.g., Bruce, 1982; Bruce et al., 1987; Hill, Schyns, & Akamatsu, 1997). A change to any of these factors leads to a reduction in identification accuracy.

In contrast, other transformations such as altering image colour (Bruce et al., 1999; Experiment 1; Davies & Thasen, 2000; Wogalter & Laughery, 1987; Experiment

1), linear perspective (Liu & Chaudhuri, 2003) or spatial resolution (Liu, Seetzen, Burton & Chaudhuri, 2003) do not always have adverse affects on performance. For instance, Bruce et al. (1999; Experiment 1) found that matching performance from arrays was better if either the target, or the referent image, or both were in monochrome, than if both were in colour. This was probably due to minor differences in perceived skin hue as the colour images were derived from different source equipment. Indeed, this result was not replicated in a second experiment suggesting that effects may be weak. Furthermore, Davies & Thasen (2000) found no difference in identification accuracy in a similar matching task comparing colour or monochrome video footage.

### 1. 6. Face recognition and matching with 'live' actors

All of the recognition and matching studies listed above utilised photographs as the target medium and yet the accused would be present in court. A court judgement in the USA specifically argued that "identification of an individual seen in a photograph is substantially less reliable than identification of an individual seen in person" (*People vs. Gould*, 1960; cited in Egan, Pittner & Goldstein, 1977; p. 200). Egan et al. suggested that 'corporeal' identification will always be more effective, as there are more available cues than are inherent in 2D 'impoverished' photographic images. A photograph can only show a single pose and it cannot replicate factors such as gait, posture, expressions, height, or weight and other elements of person recognition.

Nevertheless, conflicting results have been found when comparing identification performance to actors live in person to when they are shown in video or in photographs (E. Brown, Deffenbacher & Sturgill, 1977; Cutler & Fisher, 1990; Cutler, Fisher & Chicvara, 1989; Dent, 1977; Dent & Gray, 1975, cited in Dent 1977; Egan et al., 1977; Shepherd, Ellis & Davies, 1982). Some eyewitness memory studies have found a slight advantage in identification rates when using live targets in comparison to photographs (e.g., E. Brown et al., 1977; Cutler & Fisher, 1990; Egan et al., 1977). For instance, in a study by Egan et al. (1977)

participants viewed two target actors enacting a simulated crime through a one-way mirror. In a later identification session, they were required to select the targets from line-ups of five live actors or from two types of monochrome photograph (full length or facial frontal views) taken of the same actors. Only one of the two original actors ever appeared in the lineups, in a second lineup the target was absent. Ninety-eight percent of participants correctly identified the target actor when he was 'live'. Accuracy was lower (85%) when the target was shown in a photograph. However, 67% of participants made an incorrect false positive selection from the second lineup which did not differ across presentation modes.

Contrasting evidence for a disadvantage with live lineup targets was found by Dent for children (Dent, 1977; Dent & Stephenson, 1979; Peters, 1991) and adults (Dent & Gray, 1975, cited in Dent 1977). In both studies more participants incorrectly selected a target actor, shown live rather than from a colour photograph. Dent suggests that participants, especially children, in the live condition made hurried 'not present' selections, due to being embarrassed and nervous and less willing to closely examine the actors. Dent therefore argues that rather than a photograph advantage per se, performance would have improved if the witnesses could have viewed the live lineup through a one-way screen. However, other research has found null effects when comparing recognition of live actors to when the same actors were depicted in videos or photographs (Cutler et al., 1989; Shepherd et al, 1982).

All of the above studies examined the memory of participants. At present, only one published study appears to have been designed to examine identity matching using live actors (Kemp, Towell & Pike, 1997). Forewarned experienced supermarket cashiers were unable to correctly detect 64% of people when they presented a photo-identity card containing a 2-$cm^2$ facial photo of another person matched for facial appearance. When the distracter was simply of the same race and gender, errors were reduced, but still high at 34%. When the actors presented correct photographs of themselves, there was a relatively low false negative error rate of

7%. An explanation for this liberal acceptance criterion is that within the context of a supermarket challenging too many legitimate shoppers could result in embarrassment to the business. Furthermore, a lack of detail available in the extremely small photographs may have restricted the ability of the cashiers.

The results of these studies have forensic implications as juries can be invited to compare the resemblance of a defendant with that of a perpetrator of a crime shown in video footage. To help, they can also be provided with a photograph of the defendant taken at approximately the same time as CCTV images. One of the primary objectives of the experiments conducted in this thesis was to evaluate identification performance in this context. As such, participants viewed video footage and were required to decide if a person present at the same time was depicted in the video. Similar experiments using photographs were also conducted.

**1. 7. The effect of obscuring facial features in matching and recognition tasks**

With the increasing installation of CCTV cameras inside and outside most premises, convicted criminals have stated they would be more likely to wear a disguise if carrying out a crime (Loveday & Gill, 2003). Some disguises such as full face stocking masks or motorcycle helmets obscure all facial features. However, these would look extremely incongruous, especially if committing impulse crimes or if it was necessary to travel some distance from the scene of a crime, to avoid intensive CCTV coverage. To avoid drawing attention to activities, a more inconspicuous disguise would be likely.

Published research has consistently reported that internal facial features may be more important than external features in the matching and recognition of familiar faces. The opposite effect, or null differences are found with unfamiliar faces (Bruce et al., 1999; Clutterbuck & Johnston, 2002; 2005; Ellis, Shepherd & Davies, 1979; Henderson et al., 2001; Young, Hay, McWeeny, Flude & Ellis, 1985). Other studies have found that the familiarisation of faces is accompanied by a switch from processing primarily based on external features to one based on internal features (e.g., Clutterbuck & Johnston, 2002; 2005). For instance, Bruce et al (1999; Experiment 4) examined the matching of full-face unfamiliar video stills with photographed full-face images among arrays. When internal features (e.g., eyes, nose and mouth) in the array faces were masked by an oval, participants were 73% accurate compared to 84% when un-manipulated images were shown. However, when only the internal features were visible, accuracy was considerably reduced to 49%. This suggests that unfamiliar face matching judgments are based mainly on external features, especially hairstyle and hairline, as this is probably the most salient cue. However, photographs and stills had been taken at the same time and performance may have been reduced further in the external feature condition if images had been taken some time apart, with hairstyle altered.

No disguise would exactly obscure the face by the use of a mask over internal or external features as described above meaning these findings would not be directly relevant in a forensic setting. Most tend to partially mask either external (e.g., hats, wigs) or internal (e.g., glasses, beards) features. Nevertheless, the addition or removal of disguises between study and test has been found to reduce the accuracy with which people can be recognised by both adults and children (e.g., Diamond & Carey, 1977; Flin, 1985a; Hockley, Hemsworth & Consoli, 1999; Patterson & Baddeley, 1977, Terry, 1993; 1994). Disguise has also been found to reduce identification performance in matching studies (e.g., Henderson et al., 2001; Experiment 3). However, only two actors were recruited for the Henderson et al. (2001) study and it is unclear whether these results would be replicated with a larger pool of targets. This is explored further in Chapter 5.

**1. 8. Theoretical explanations for the unfamiliar face processing disadvantage**

Theoretical explanations for the differences in processing of unfamiliar and familiar faces have been proposed by Bruce and others (Bruce, 1982; Bruce & Young, 1986; Burton, Bruce & Hancock, 1999). These models implicate separate functions, with the recognition of familiar people involving the development of view-invariant structural abstract representations stored in long term memory. Once these representations are formed, exposure to a known face allows for high levels of facial recognition, even from degraded images, or changes in angle, lighting or facial expression. The finding that familiar faces are recognised more effectively using internal features reflects the relative stability of the configural relationship between features, allowing recognition to be based on matching to the permanent representations stored within the cognitive system.

The identification of unfamiliar people is proposed to be governed by less flexible pictorial elements, or episodic representations, involving the processing of viewpoint, expression and lighting-specific codes. If exactly the same images are used in study and test, recognition memory for unfamiliar faces can be extremely good as memory for these pictorial elements will be high (Bruce, 1982). However, if lighting, expression or pose is altered, the codes do not match and recognition is poor (e.g., Bruce, 1982; Bruce et al., 1987; Hill et al., 1997). For this reason, even when there are no demands on memory, people are less efficient at extrapolating from one view of an unfamiliar face to another (Bruce, 1982; Bruce et al., 1999; Henderson et al, 1999; Young et al, 1985).

Bruce and Young (1986) argue that representations of newly-encountered people take time to develop in order to mediate accurate recognition. Evidence is admissible in court if a witness not previously familiar to a defendant spends substantial time examining footage and subsequently identifies a suspect (e.g., Attorney General's Reference, No. 2 of 2002, 2003; *R v Clare and Peach*, 1995). However, it is unclear how much and what type of exposure to an unfamiliar face is required before it acquires the same representational properties as a familiar face. Bruce et al. (2001; Experiment 2) found that viewing moving unfamiliar faces for up to a minute did not improve recognition accuracy in a subsequent matching test. However, in a later experiment (Experiment 3), they asked pairs of participants to "chat about the faces between yourselves as you watch the video" (p. 215). This initial socialisation process resulted in an increased hit rate and reduced false alarm rate in a following task in which participants matched faces from arrays. However, it is unclear whether improved accuracy was associated with social factors per se, or whether discussion induced a deeper depth of processing found in other domains to be associated with improved memory performance (e.g., Craik & Lockhart, 1972). Other published research has investigated how much experience with a face is required before performance on other perceptual tasks is altered (e.g., Bonner, Burton & Bruce, 2003; Clutterbuck & Johnston, 2004; 2005; O'Donnell & Bruce, 2001). The results of these studies do not directly impact on the procedures involved in court cases such as *R v Clare and Peach* (1995). However, they do provide an indication that it appears possible to be familarised to an individual from viewing extremely brief facial video clips. Nonetheless, the formation of these 'immature' representations may be relatively quick, but it is unclear whether they are stable.

**1. 9. The effect of distinctiveness and perceived similarity on face processing**

A large body of research has found that faces rated as distinctive are recognised more efficiently than those rated as typical. This results in lower false alarm rates when distracters and higher hit rates when targets, as well as in speeded response times. The effect has been found with highly familiar and celebrity faces (Valentine & Bruce, 1986a), and when participants are tested on their recognition of previously unknown faces (e.g., Light, Kayra-Stuart & Hollander, 1979; Valentine, 1991; Valentine & Bruce, 1986b). In contrast, studies presenting sequences of un-manipulated and jumbled faces have found that typical faces are categorised as faces faster than distinctive faces (e.g., Valentine & Bruce, 1986b).

These effects are explained by Valentine's (1991) *face space* model of face recognition in which faces are represented in a multidimensional face space. Each point in this space represents a previously encountered face. Although the number of dimensions is not specified, Valentine suggests that this will be equal to the number of properties on which faces can be differentiated. The origin of the space represents the central tendency of the population of previously encountered faces with points representing faces clustered around this location. The status of a face (distinctive or typical) is determined by its position within the space. Typical faces cluster near to the origin, and by definition will tend to possess similar attributes on some dimensions, whereas distinctive faces will be more dispersed. The finding that distinctiveness facilitates recognition is explained as being due to efficient access to those faces as they are less likely to be located close to, and confused with other exemplars in face space. In contrast, typical faces being closer exemplars to the central tendency will be classified more quickly as a face when presented among jumbled faces.

The issue of facial similarity and distinctiveness has not been specifically addressed in published studies examining simultaneous face matching. In an experimental context, it would be unlikely for participants to incorrectly match a distracter that was rated as dissimilar to a target. Indeed, most research has used various methods to preliminarily match targets with distracters to produce perceptually similar facial arrays (e.g., Bruce et al., 1999). However, the face space model has been successfully applied to the recognition of facial and bodily movement, especially if this is artificially caricatured (Hill & Pollick, 2000; Lander & Chuang, 2005). It has also been applied to the difficulties people often have with distinguishing between the faces of people of other races and from other demographic groups (e.g., Chiroro & Valentine, 1995; Valentine, 1991).

### 1. 10. The effect of movement in person identification

Moving CCTV images are acquired at a rate of 25 (Europe) or 29 (USA) frames per second. Most operators can also select continuous recording from a specific camera if they believe an incident is occurring, and this may happen automatically if movement detection equipment is installed. However, due to storage costs, multiple stills are often intermittently recorded. The acquisition of still images will have implications on detection rates as the equipment may ‘miss’ an ideal viewpoint of a moving target’s face. Furthermore, some commentators have suggested that movement allows extrapolation of 3-D structural information about people, not available from single views (Schiff, Banka & de Bordes Galai, 1986). Indeed, movement of facial expressions and whole faces (e.g., nodding) is advantageous for familiar person recognition when images are highly degraded (e.g., photographic negatives, Knight & Johnston, 1997; inverted, pixilated or blurred, Lander, Bruce & Hill, 1999). This advantage is particularly found with people possessing highly distinctive idiosyncratic movements (Lander & Chuang, 2005).

Furthermore, when minimal cues are provided, for instance, using point-light displays in which only points of light fixed to actors' bodies are visible, participants tend to be able to identify highly familiar people from the movements alone at better than chance levels (Cutting & Kozlowski, 1977). Similar effects are found from light displays of facial features (Bruce & Valentine, 1988). Cutting and Kozlowski (1977) note that the most accurate observers tended to report that decisions were based on specific dynamic features such as walking speed, bounciness and rhythm, possibly an indication of individual differences in the awareness and attention to these features. Indeed, Loula, Prasad, Harber & Shiffrar (2005) found that participants were better at identifying themselves or a friend in point-light displays when performing more unusual actions such as boxing or table tennis than when walking or running.

In contrast, conflicting results have been found when examining the effect of movement in the identification of unfamiliar people. For instance, Shepherd et al. (1982) and Christie and Bruce (1998) found no recognition advantage for movement over stills, even when available information was matched. However, Schiff et al. (1986) found that after viewing a film of a staged robbery, dynamic mug shots showing faces turning 180° from right-to-left elicited higher recognition rates than multiple stills. In addition, Cutler, Penrod and Martens (1987) found that eyewitness recognition accuracy was better if lineups of moving actors were shown rather than facial photographs, especially if the distracters in the lineup were highly similar in appearance to the target. Thus it appears that movement may help to disambiguate between homogenous individuals under difficult identification conditions.

However, there are some circumstances in which movement appears to have an adverse effect on unfamiliar face identification. Liu et al. (2003a: Experiments 1 & 2) presented either moving or still low resolution poor-quality video footage alongside either low or high resolution photographs in a design that examined both recognition memory and simultaneous matching ability. If the video footage and the referent photograph were of equally poor resolution, movement in the video was found to be detrimental to recognition but not to matching. In contrast, if the still referent image was of high quality, movement was not detrimental to either task.

All the experiments reported in this thesis used high quality video footage and photographs to ensure that the effects found by Liu et al. (2003a) would not be replicated. Furthermore, in a courtroom juries would be shown the entire relevant video footage captured by CCTV, so to simulate this situation, moving images were always presented.

**1. 11. Demographic effects in person identification**

There is evidence of an own-demographic group advantage in face recognition across a number of domains. The most regularly studied has been the own-race effect in which members of one's own race are recognised more effectively than those from other races (e.g., Meissner & Brigham, 2001; Shapiro & Penrod, 1986). Conversely, other-race faces are classified as faces faster in race definition tasks (Valentine & Endo, 1992). A large component of this effect may be due to a lack of contact or interest in members of other racial groups, reducing ability to select relevant discriminating features (Chiroro & Valentine, 1995). In a meta-analysis of the literature, Meissner and Brigham (2001) demonstrated that white Caucasian participants are particularly susceptible, resulting in an increase in false alarms, when identifying faces from other races.

There is also evidence of an own-gender advantage, especially with female faces (Lewin & Herlitz, 2002; Shapiro & Penrod, 1986; Terry, 1993; Wright & Sladden, 2003). Females also tend to select more incorrect foils than males in real lineups

(Valentine et al., 2003). Wright and Sladden (2003) note that a large component of the own-gender effect is that the encoding of own-gender hairstyles is more efficient, leading to a greater susceptibility to recognition failures if the hairstyle of someone from the opposite gender is changed or disguised.

Eyewitness studies have also found that age negatively correlates with identification, with a substantial decrement in performance above 50-years of age (O'Rourke, Penrod, Cutler & Stuve, 1989) leading to an increase in false alarms (Smith & Winograd, 1978). Older eyewitnesses also tend to make less positive identifications than younger adults from real lineups (Valentine et al., 2003). This may be in part due to a general cognitive decline. Indeed, Searcy, Bartlett and Memon (1999) found that older adults, screened to ensure that they did not fail a perceptual discrimination task performed as accurately at face recognition as younger adults. However, detrimental performance may actually be the result of an own-age advantage in face identification (Chung & Thomson, 1995; Fulton & Bartlett, 1991; George & Hole, 1995; Wright & Stroud, 2002). For instance, George and Hole (1995) found that estimations of age are more accurate with people of the same age. Furthermore, Wright and Stroud (2002) found that young adults aged 18 – 25 and older adults aged 35 – 55 were better at selecting a previously seen target photograph from a lineup when that target was of their own age group.

These demographic effects have been interpreted in terms of Valentine's (1991) face space model, in that the faces of other groups are encoded in multidimensional space further from the central tendency in a similar manner to distinctive faces. However, other-group faces are clustered around a secondary 'satellite' norm. Therefore, although possessing distinctive characteristics in terms of their location in face space, they are less distinguishable from others in the same cluster. Increased exposure to other-demographic group faces is believed to improve the representations of the important features that differentiate between these faces, so that they are more evenly spread throughout face space thereby facilitating recognition performance.

The findings of deficits in the processing of specific demographic groups are of concern. No published studies appear to have specifically compared older and younger adults, or the ability to distinguish between people of other races on a face matching task. However, in the UK most crimes are committed by younger adults and a disproportionate number of these are members of minority ethnic groups. With an ageing population, randomly selected juries are more likely to be made up of adults of up to 70-years of age. In terms of evidence presented in court, there is a potential for error if predominately older jurors are invited to compare the resemblance of a young defendant to a perpetrator of a crime depicted in CCTV footage.

**1. 12. Face processing by children**

Even if facial stimuli are matched in age to participants, children tend to be inferior on most face processing tasks. Indeed, children's memory for faces improves from 2-years until late adolescence (Blaney & Winograd, 1978; Bruce, Campbell, Doherty-Sneddon, et al., 2000; Carey, Diamond & Woods, 1980; Davies, 1993; 1996b; Diamond, Carey & Back, 1983; Flin, 1980; 1985a) with evidence of a leveling in performance from the ages of 9- to 13-years. Some studies have found a performance dip at the same time as puberty, being most pronounced at 12-years of age, with accuracy then improving until adult proficiency is attained (e.g., Carey et al., 1980; Diamond et al., 1983; Flin, 1980; 1985a; Soppe, 1986).

Conflicting results have been found when examining the maturation of face recognition in children using eyewitness paradigms. Identification from photo line-ups reaches adult levels by approximately six years of age, but only when the target is present (e.g., Goodman & Read, 1986; J.F. Parker, Haverfield & Baker-Thomas, 1986; J.F. Parker & Carranza, 1989). There is also evidence of a response bias in that children make more selections regardless of accuracy, leading to an increase in false positive identifications (Lindsay, Pozzulo, Craig, Lee & Corber, 1997; J.F. Parker & Carranza, 1989; J.F. Parker & Ryan, 1993). The authors suggest that this may partly be due to children believing that the task would be meaningless unless the target was present and that children guess in order to please the experimenter. In contrast, adults, being more restrained, tend to make fewer selections, increasing target present errors.

There are also qualitative differences in the processing of faces by adults and children. In adults, the internal features of faces are of more importance than the external features in the recognition and matching of *familiar* faces (Ellis et al., 1979; Young et al., 1985). In contrast, younger children tend to show an external feature advantage for face recognition and matching (Bonner & Burton, 2004; Newcombe & Lie, 1995). There is conflicting evidence as to exactly what age this shift occurs. However, Campbell, Coleman, Walker et al. (1999) found evidence that the inner face advantage may not fully develop until the age of 15. In contrast, matching studies have found evidence of a highly-familiar internal feature facial advantage in children of 7 and above (Bonner & Burton, 2004).

An external *unfamiliar* face advantage has been found in adults and older children at both matching and recognition tasks (Bonner & Burton, 2004; Bruce et al., 1999; Ellis et al., 1979; Want, Pascalis, Coleman & Blades, 2003; Young et al, 1985), although for children of 7 – 8 years-of-age, matching of unfamiliar faces appears to be equally effective using either internal or external features (Bonner & Burton, 2004). However, Newcombe and Lie (1995) demonstrated that children from 4 - 6 years were more accurate at matching both unfamiliar and experimentally-familiarised faces using external features. This suggests that both adults and children utilise external features when matching or recognising unfamiliar people, but for adults and older children internal features are more important when identifying familiar people.

No published studies appear to have directly compared face matching skills in children and adults. However, Ellis (1992) reported improved performance from 3 to 8-years of age on a matching task involving transformations of expression, pose and facial paraphernalia (e.g. glasses). There was also a trend for an improvement from 8 to 11-years of age although accuracy was close to ceiling in this group. Furthermore, Bruce et al. (2000) using an identity-matching two-alternative forced-choice task in which facial pose was altered either in targets or in distracters found that performance reached ceiling at 7-years of age if faces were dissimilar in appearance. For those rated as similar in appearance, a clear developmental progression was demonstrated with accuracy not reaching 100% even in 10-year-olds, the oldest children tested. If external features such as hairstyle were masked, performance was reduced in all age groups. Nevertheless, in both of these studies (Bruce et al, 2000; Ellis, 1992), performance was at or near ceiling suggesting that the task may have been relatively easy and therefore it is unclear whether the oldest children were performing at adult levels, or that face matching skills reach their peak at a later age.

Carey et al. (1980; Experiment 3) suggest that performance at face matching is similar to that of recognition, in that there is a developmental dip at approximately 10-years of age. However, whereas with recognition tasks, improvement appears to continue following a brief hiatus of one or two years, in matching tasks the plateau or reduction remains until 14-years of age, with a subsequent improvement until adult levels are reached by the age of 16.

The forensic importance of measuring face matching skills in children relates to their reliability as eyewitnesses. If they are worse than adults at matching two images of the same person, or at discriminating between two images of different people, then it would be legitimate to question their competence at correctly identifying an unfamiliar adult after witnessing a crime.

**1. 13. Thesis overview**

The main aim of this thesis was to perform a systematic evaluation of the manner in which CCTV evidence may be used in different circumstances by the criminal justice system. As such, the primary focus was to assess three of the four recommendations outlined in the Attorney General's reference as to the admissibility of CCTV evidence for identification purposes in courts (Attorney General's Reference, No. 2 of 2002, 2003). The first recommendation, which was not tested in depth, concerned the attendance of witnesses previously familiar with the accused providing opinion evidence as to identity. Studies investigating highly familiar person recognition even in poor-quality video footage have found that this type of identification is normally extremely accurate. Therefore, such testimony would be expected to be reliable.

The second recommendation was that if CCTV images are 'sufficiently clear', jurors may be invited to match the image with the defendant to form their own opinion as to whether it is the same person or not. For less serious crimes this may also be required of Justices of the Peace (JP) in a magistrate's court. In addition, regardless of whether they are actually asked to perform this task, an individual juror or magistrate is potentially liable to make this type of judgement if CCTV evidence is submitted. In these circumstances all would be unfamiliar with the defendant and in contrast to studies examining familiar face recognition, unfamiliar face matching even from high-quality footage is often unreliable (e.g., Bruce et al., 1999; Bruce et al., 2001; Henderson, Bruce & Burton, 2001).

In court the defendant will always be present in person, therefore a series of experiments were conducted using a forensically and ecologically-valid methodology to replicate the task that might be required of a jury. Single-item identity-verification designs were conducted in which participants made judgements as to whether an actor physically present in person was simultaneously depicted in video footage. Different quality footage was employed and the time between video capture and identification session was also varied, to simulate circumstances that might occur in a criminal investigation. Across all experiments, target present and target absent trials were conducted, in which the individual on video was replaced by a distracter as might occur if an innocent suspect was arrested. The second recommendation from the Attorney General's reference also proposed that an undisputed contemporary photograph of the accused, taken at approximately the time of the crime should be available to the jury. Therefore, further experiments were conducted using photographs as stimuli. In one, performance was directly compared to when the actors depicted in the photographs attended identification sessions.

Chapter 3 provides the full specifications of the stimuli such as the video footage and photographs used in the experiments reported in this thesis. Details of the actors and the results of a series of pilot studies are also described. These were conducted to ensure that conditions would replicate those found in real forensic scenarios.

Two experiments are reported in Chapter 4 using video footage designed to simulate that which might be obtained by a typical open-street operated-controlled CCTV system. In Experiment 4.1 participants were required to match a target shown in the footage with a photograph from within an array of six. The task was simplified in Experiment 4.2 with a single-item identification-verification design. In both experiments, performance across age groups was also examined.

Two experiments using the single-item identification-verification design are also reported in Chapter 5. These were designed to investigate adult face matching when targets were wearing a disguise. In Experiment 5.1 videos showing the actors in three different disguise conditions were used, with referent photographs depicting the actors in no disguise. In contrast, in Experiment 5.2, disguise was manipulated in the photographs instead.

The same design was used in three experiments reported in Chapter 6 in which participants made judgements as to whether physically present actors were shown in simultaneously presented video footage. The footage from Chapter 4 was used in Experiment 6.1 which at the time was three weeks old. The same design and footage was used in Experiment 6.2, although in some trials the actors were shown in disguise. The identification sessions took place a year after filming to imitate events that commonly arise within the criminal justice system. Half of the participants were informed of this, allowing assessment of whether responses on this primarily perceptual task were influenced from knowing that appearance might have changed. In Experiment 6.3 the performance of children was compared with adults using the same materials as in Experiment 6.2.

One experiment is reported in Chapter 7, using a single-item identification-verification design. High-quality close-up facial video footage was obtained and matching studies were initiated with identification sessions taking place either a few minutes or a week after video capture. A direct comparison of matching to actors physically present and matching to the same actors shown in photographs was also conducted.

The third recommendation to the Attorney General was that opinion evidence should be admissible if a witness previously unfamiliar to a perpetrator views video footage until they claim to have familiarised themselves with individuals shown in the images (Attorney General's Reference, No. 2 of 2002, 2003). The experiment reported in Chapter 8 examined identification accuracy in this context. Participants extensively viewed video images over a period of approximately a week. Performance at a final matching task was compared with others who had viewed the footage only a few times as expected for a jury.

In Chapter 9 the deliberation processes of participants acting the part of a jury were examined. Each group of twelve participants took part in two successive mock jury 'trials' in which the core evidence was surveillance video footage. In both cases they were invited to compare the resemblance of the 'defendant' shown in a photograph with the 'offender' shown in CCTV footage. Private juror-level responses were collected and compared with public jury-level individual and group voting preferences. This allowed an examination of whether juries reminded of the 'beyond reasonable doubt' standard of proof would render a verdict based on CCTV evidence alone.

The final recommendation within the Attorney General's reference (Attorney General's Reference, No. 2 of 2002, 2003) relates to evidence provided by experts in facial structure. Techniques used by expert witnesses presenting evidence on facial identification are discussed in Chapter 10. In many cases, these entail the use of facial mapping to provide an indication of the probability of a match with the accused. Custom software was designed to aid in the identification of specific facial landmarks in photographs and to automatically provide a database of both the physical and angular distance between these points in two-dimensional space. This necessitated the collection of a database of facial images and the application of a number of different statistical analyses. The primary aim was to examine whether a specific face on one photograph could be correctly identified as the same person in a second photograph.

The thesis commences in Chapter 2 with a review of literature describing the social effects of CCTV, and the development of automatic recognition systems. Most research on the crime reduction effects has been conducted in the UK, probably due to the pioneering establishment of comprehensive video surveillance systems in this country. Therefore, it is likely that the effectiveness of CCTV in different research contexts will impact on its use elsewhere in the world. Knowledge of the history of CCTV and its success in meeting its publicised aims can inform potential stakeholders as to the conditions under which implementation is likely to be most successful. This has particular implications for the use of CCTV for identification purposes as technological advances have often been designed for this purpose.

## Chapter 2: The social impact of CCTV and computerised recognition systems

### 2. 0. Introduction

In terms of the number of CCTV cameras, by the end of the twentieth century the UK was the leading exponent in the world (Norris & Armstrong, 1999). However, with increased investment in many countries, it was predicted it will soon be overtaken, in terms of both density and in the actual number (Norris et al., 2004). The UK has pioneered research into CCTV, especially when examining its effectiveness as an identification tool. The results of these studies have to be understood in terms of the contemporary implementation objectives and technological capabilities. Indeed, the history, the prevalence, the quality and the management of systems in the UK all have a direct impact in terms of the primary topic of this thesis. For instance, two separate studies examining whether the introduction of town centre CCTV reduced crime found that to some extent this was dependent on the likelihood of operators recognising potential offenders (Ditton & Short, 1998; Ditton, Short, Phillips, Norris & Armstrong, 1999). Furthermore, convicted offenders have stated that they would modify their behaviour, to reduce the probability of identification, if image quality and scheme management were to improve (Gill & Loveday, 2003a).

### 2. 1. The history of CCTV in the UK

The earliest proposal for a CCTV system in the UK was to manage crowd-control at the Royal Wedding in 1947. However, the first actual camera was a one-man traffic light operation in Durham in 1956. The Metropolitan Police initially used portable CCTV to monitor demonstrations and public events. This included a visit to London by the Thai royal family in 1961 and a number of anti-Vietnam war protests later that decade. By the end of the 1960's there were 67 cameras nationwide, operated by 14 different police forces (Williams, 2003). The success of these schemes led to the permanent installation of cameras in the political centre of the capital in 1969 and there are currently over 260 cameras in Parliament Square alone (POST, 2002).

With the invention of videotape in the 1960's, recording and storing images in a central control room became possible. This led to the development of the first retail systems by the company *Photoscan* in 1967. CCTV was introduced on the overland railway system at Dagenham in 1965 (Williams, 2003), London Underground in Holborn in 1961 (McCahill & Norris, 2003b), and by 2006, there were expected to be 9,000 cameras throughout the underground network (Hogan, 2003). London's buses, bus lanes, traffic control systems and airports are also covered by thousands of surveillance cameras (McCahill & Norris, 2003b), along with 6,000 speed cameras on roads across the UK (POST, 2004). With the introduction of the Traffic Congestion Charging Zone in central London in 2003, 700 further cameras were installed, designed to read number plates and to photograph the driver of every vehicle (McCahill & Norris, 2002c).

The first permanent local authority open-street systems were introduced in Bournemouth in August 1985 (Bannister, Fyfe & Kearns, 1998) and by 1991 there were approximately 10 city centre schemes in the UK. McCahill and Norris (2003a) suggest that an important catalyst for large scale introduction was due to fear of IRA activity. However, the doubling of crime rates between 1979 and 1992 probably facilitated this process, as did the associated positive publicity from the release of CCTV stills depicting the high profile abduction of Jamie Bulger by his child murderers in February 1993 (Norris & Armstrong, 1999). Indeed, the following year the Home Office announced an initial City Challenge Competition to allocate £2 million for open-street CCTV, if finance could be matched by local businesses. Bids were received for 480 schemes and further initiatives were announced, so that between 1998 and 2002, £170 million was allocated by the Government to fund 1,300 systems, across 78% of local authorities (POST, 2002; Webster, 2004). However, Home Office initiatives do not include other publicly funded schemes such as schools, hospitals or universities so the actual Government outlay will have been much higher. Indeed, Norris et al. (2004) estimated that with the inclusion of privately-funded schemes, between 1994 and 2004, £4 - 5 billion had financed the installation and operation of CCTV in the UK. Furthermore, one of fastest growing sectors is home security, with one company announcing that 90% of its sales were for the domestic market (BBC News, 7 April, 2005). The growth of this sector has led to estimates that there will be 25 million cameras in the UK by the end of 2007 (Honore, 2004).

**2. 2. The operation of local authority CCTV schemes**

Most systems designed to monitor open public spaces are operated under the auspices of local government and involve networked cameras connected to an observation centre, allowing monitoring, recording and storage of footage (Webster, 2004). Most current equipment is of low specification so that images of people shown in footage are often unclear. The police in the UK do have image enhancing equipment, although there is limited scope with analogue film; which is the medium normally used in court. With the emergence of digital systems some of these problems may be overcome, as analogue systems already in place are likely to be replaced. Indeed, some digital systems can zoom into specific areas within a visual scene, whilst retaining the main image, either in real time or during post-event analysis (Verdant Technologies, 2004).

However, the algorithms designed to enhance images, perhaps for identification purposes may create inaccuracies in digital files. It is extremely difficult to detect alterations to files, such as the addition of people or objects to images. Frames can also easily be removed or added. As such, the authenticity of evidence could be questioned in court. The House of Lords Select Committee on Science and Technology ($5^{th}$ Report, 1997/1998) noted that it also would be relatively simple for a criminal to manipulate digital footage to create an alibi. Bull (2003) argues that as long as there is a detailed audit trail, digital evidence should have high probative value. Various encryption techniques are available, which produce the digital equivalent of a watermark. This is automatically added at image capture and can only be accessed with a decryption code. However, digital watermarks can subtlety affect algorithms that are designed to identify individuals, although this would not be apparent to a human operator. Pramateftakis, Oelbaum and Diepold (2004) suggest that the use of digital signatures generated by individual cameras and embedded in the bit stream may provide a solution. Nevertheless, the House of Lords Committee also noted that the cost of challenging a complex audit trail might be prohibitive. Due to these unresolved issues, Murphy (1999) believes that the use of digital video evidence is likely to be challenged over the next few years, predicting that these challenges will lead to the emergence of "well developed principles of admissibility" (p. 401).

Bull (2003) suggests that a primary focus for the future will be the integration of many different systems in order to facilitate the rapid "capture and exchange of high-quality multimedia information" (p. 142) nationally from multiple sources. However, it appears that the public finance required to update all systems to digital in the UK is unlikely to be available in the near future. The final Home Office CCTV initiative was conducted in 2002 and no large scale plans were put in place for central Government to fund financially prohibitive replacements (Irving, 2005). Indeed, only 15 of the 6,000 cameras in operation on the London Underground are digital (*Independent*, 23 August, 2005). However, in less saturated markets around the world, the establishment of high resolution digital systems is probable and this may lead to criticism, if the UK is perceived in the future to have second-rate systems.

**2. 3. Automatic recognition systems**

Many computer engineers have been engaged on the development of algorithmic pattern recognition systems designed to identify faces. These would perform two functions. One is for verification of an individual, for instance, to ensure authorised access to a secure building. The second is for identification purposes, so that an alarm would be triggered if a target individual whose face is on a database enters a monitored area. In both cases, a human facial still or moving image is extracted from the background film. This is then transformed into an abstract representation or *biometric*. This unique individual identifier is then compared to a gallery of facial images. High power systems are designed to monitor a series of stills until a 'best' image is presented and a probability for a match is generated. This type of system is already in regular use along with other biometrics such as fingerprints or DNA. However, facial images can be acquired without any active participation, consent, or even knowledge of a target.

Some commentators have predicted that when perfected, automatic systems will perform as efficiently as cameras designed to read car number plates (e.g., Norris & Armstrong, 1999). If so, in the UK since 2003 all vehicles entering the London Congestion Charge Zone are automatically identified if included on the Driver and Vehicle Licensing Agency (DVLA) database (McCahill & Norris, 2002c) and the Government recently announced plans for a similar scheme on all major highways (Knight, 2006). The details of 17.9 million registration documents were amended or added to this database in 2003 (DVLA, 2004), and the system itself has an extremely low failure rate. However, the algorithms required to match two faces are far more complex than those required for reading standardised number plate letter and numeral shapes.

Comprehensive digital facial databases are already established in the UK. The police routinely photograph everyone charged with an offence (POST, 2001) and all passport and driver license applicants already have their photos placed on a digital database. In 2003 – 2004, 6.5 million driving licenses (DVLA, 2004) and 6.1 million passports (UK Passport Service, 2004) were issued. With the proposed introduction of a national identity card, it is likely that a national face database will also be instigated. Indeed, the Government has stated that biometric data, including face images will be compiled for every individual in the UK. During the parliamentary debate as to the implementation of the national identity card, the Police Information Technology Organisation (PITO) announced plans for a Facial Images National Database (FIND) linked to criminal records in conjunction with facial recognition technology (Ranger, 2006).

Whilst the detailed algorithms for face identification utilised by commercial companies are strictly confidential, a large body of academic research has been published on this issue (e.g., Brunelli & Poggio, 1993; Kirby & Sirovich, 1990; Turk & Pentland, 1991). However, the technology involved in just detecting a face and its landmarks from a background scene is complex and requires an extremely high computational load (Feraud, Bernier, Viallet & Collobert, 2001; Hjelmas & Low, 2001). Furthermore, faces can be partially occluded, for instance, if in a crowd, or in shadow, meaning that appearance is constantly being altered. Due to these inherent problems, Socolinsky, Selinger and Neuheisel (2003) suggest that using infra-red facial thermogram systems may prove more successful, as they measure the unchanging pattern of arteries and veins underneath the skin. However, this would require the financially prohibitive conversion of all current systems and databases.

According to Brunelli and Poggio (1993), systems can be classified into two categories; geometric feature-based, and global template-based techniques, although some have a degree of overlap (e.g., Takacs, 1998). Geometric feature-based techniques extract and measure discrete local features, employing statistical pattern recognition methods for retrieval and identification (e.g., Brunelli and Poggio, 1993; Wiskott, Fellous, Kruger, & von der Malsburg, 1997). These systems are related to photo-anthropometric analyses, in that the relative placement and distance of internal facial landmarks are calculated and compared. In contrast, template-based techniques use global representations, applying principal components analysis (PCA) to the different intensities of image pixels in photographs (e.g., Kirby & Sirovich, 1990; Turk & Pentland, 1991). Heisele, Ho, Wu and Poggio (2003) compared two globally-based systems with a feature-based system finding that face identification performance was more effective in the latter, especially when facial pose differed, as global patterns of pixel intensity are particularly susceptible to viewpoint changes. However, Hancock, Bruce and Burton (1998) found the PCA-based systems operate more closely to human perception on a number of tasks, and when viewpoint was matched were more effective than feature-based software.

A series of large scale US Government sponsored independent automatic system tests have been undertaken using a large database of faces (Face Recognition Technology; FERET). The most recent tested the verification and identification performance of ten different commercial systems against a database of 121,589 facial images of 37,437 different individuals (Facial Recognition Vendor Test (FRVT); P.J. Phillips et al., 2003). To examine performance, different novel probe face images were entered into the systems to match with the database, a test analogous to issuing an old-fashioned 'wanted poster'. The most successful system in a controlled indoor environment was found to have, at a false alarm rate of 1%, a 73% correct verification rate if only the best match was examined. This rose to 82% if the top ten ranked faces were accepted. In an outdoor environment, verification at the same false alarm rate (1%) was reduced to 50%, mainly due to the algorithms coding facial shadowing effects as actual facial features. Time lapses between image capture of the same individual further reduced performance by approximately 5% per annum, meaning that databases would need to be constantly updated.

From these figures Introna and Wood (2004) calculate that if a UK mugshot database contained a similar number of faces as the current finger print database (5.5 million), identification performance would be "approximately 55% in ideal conditions and as low as 32% in less than ideal conditions" (p. 189). It would be possible to accept a higher false alarm rate, consequently increasing the number of correct identifications. However, public confidence in a system would be undermined if too many individuals were constantly over-scrutinised. Those most at risk of failing a verification test would also be those possessing an algorithm-derived 'typical' or standard face. This could have political ramifications. Indeed, Norris and Armstrong (1999) describe the case of two football fans, erroneously placed on a database of suspected hooligans and identified, arrested and deported by

Belgian police. The authors suggest that their 'real selves' had less authority than their database classification.

Field trials have also been conducted. For instance, Newham Council in London introduced a scheme linking 140 cameras in 12 shopping centers. Images were checked against a database of 100 known criminals. During August 2001, 527,000 separate faces were detected and 90 were positively matched against the database. Installation resulted in a 34% reduction in street robbery. However, there were no arrests and the positive crime effects were probably due to the system acting as a deterrent, as all database targets were pre-warned of the scheme (POST, 2001).

Studies comparing human and computer face identification ability have also been conducted. For instance, A.M. Burton et al. (2001) directly compared a computer PCA-based system with the performance of humans participating in the Bruce et al. (1999) face matching studies. In that experiment, a high-quality close-up full-face video still was simultaneously presented alongside target present arrays of 10 high-quality photographs. Whereas humans failed to correctly identify 24% of targets from the series of 20 arrays, the error rate with the most successful PCA technique was only 6%. In a second study, pose was changed (three-quarters) in the video still. In this condition, humans made slightly more errors than in the first unchanged-pose study (29%). However, using the best PCA-based system, the error rate was far higher (40%). A further indication that global-pattern algorithms are most effective when photographic viewpoint is matched.

As well as investigations into face recognition, parallel research has been undertaken on other automatic identification systems such as ear (e.g., Hurley, Nixon & Carter, 2005; Moenssens 1999) and gait recognition (e.g., Nixon & Carter, 2004). Both types of data can be acquired covertly and there is evidence that each provides a unique human signature. However, there has been a recent successful challenge in the courts against the use of ear-print evidence, for its lack of reliability (Woffinden, 2004). Furthermore, comprehensive gait and ear databases, similar to that of fingerprints would need to be established for forensic or verification purposes. Other innovations are being developed. For instance, Intelligent Scene Monitoring software designed to detect individual or unusual suspicious behaviour amongst crowds potentially numbering thousands has been tested on the London Underground (Silicon.com, 29 March, 2004).

Introna and Wood (2004) suggest that following the September 11th 2001 terrorist attacks in New York, investment in the biometric industry has rapidly expanded. Annual global expenditure is expected to rise from $719 million in 2003 to $4.6 billion in 2008 with the face-specific biometric market rising from $50 million to $802 million within the same time frame (Sarker, 2004). A recent report to the European Commission predicted that with the introduction of 3-D analysis, pre-processing of higher resolution images and automatic expression identification techniques, there will be a significant improvement in face recognition accuracy, particularly in environmentally natural conditions (European Commission, 2005).

**2. 4. The implementation of CCTV schemes**

CCTV installation has traditionally been marketed as a crime prevention measure, designed to fulfill two basic purposes. The first is to act as a deterrent to criminal activity; the second, to aid in the identification of suspects and to provide forensic evidence when a crime has been committed. Doubts have been expressed as to its success as a deterrent (e.g., Gill & Loveday, 2003a), and a number of studies have also shown that the identification of unfamiliar people in CCTV images may be unreliable (e.g., A.M. Burton et al., 1999; Bruce, et al. 2001). However, some commentators have suggested that the introduction of a local authority CCTV system may serve different functions to the various stakeholders involved (Ditton & Short, 1998; Gill et al., 2003; Reeve, 1998). Indeed, Reeve (1998) identified a number of business and political aims. Primarily these are to cut crime, anti-social behaviour and vandalism, to reduce the fear of crime, and to attract more people to an area. Associated objectives may be to manage traffic and parking, and to create a pleasant environment, conducive to increased retail and business activity. Therefore it is apparent that the success of a scheme should perhaps not only be measured entirely in terms of crime reduction, but also how it provides a wider benefit. This can be examined by measuring whether the support found in advance of CCTV introduction remains in place over time.

**2. 5. Public and Political Support**

Generally, the public in the UK is enthusiastic in advance of the local introduction of CCTV (Ditton, 2000; Ditton & Short, 1998; Gill et al., 2003; 2005; Honess & Charman, 1992; Winge & Knutsson, 2003). Gill et al. (2003) found that between 77% and 94% of 4,400 residents in nine different districts covered by seven local authority schemes were in favor of installation. Most respondents believed that crime would be reduced; often stating that due to a perceived increase in safety, they would enter previously avoided areas. In contrast, only 17% of the respondents believed that CCTV would be an invasion of their privacy. Follow up surveys have found that CCTV retains public support some time after introduction, although the percentage is often down (Ditton, 2000).

Nevertheless, public support for the installation of widespread CCTV in other countries is much lower. For instance, in a survey of Berlin residents, Helten and Fischer (2004) found strong support for cameras for banks and subways. However, there was more opposition than in the UK to placement in high streets and residential building lobbies. In these countries, the belief in the importance of privacy and of a less intrusive state was seen as a paramount issue. These factors are discussed later in this chapter.

Furthermore, whilst CCTV-surveilled areas are seen as safer, fear of crime is not necessarily reduced (Ditton, 2000; Gill et al., 2005). This may be due to a perception that the installation of cameras is associated with areas of greater risk. In addition, people feel they are less accountable for the welfare of others, as areas are monitored by those with responsibility, generating an increased sense of personal isolation (Ditton, 2000). Ditton (1998) suggests that the high level of public approval in surveys is related to preliminary leading questions that heighten awareness of the fear of crime. When initial questions highlight negative connotations, such as potential civil rights and privacy issues, or abuses of access to images, approval tends to be much lower. Furthermore, Norris and Armstrong (1999) suggest that police and political announcements have reported success in reducing crime when most independent analyses confirm much lower rates. Studies finding negative or null effects tend to be publicly criticised as being flawed or unrepresentative.

Indeed, it is perhaps unsurprising that the public in the UK tend to be supportive of CCTV use. Most of the information they receive comes from the media which tends to focus on specific cases with successful outcomes (Webster, 2004). This occurred following the rapid identification of the London suicide bombers in July 2005. All news reports in the UK contained these images (e.g., BBC News, 20 July, 2005) and across the world, news agencies commented that the high numbers of cameras in the city aided the police, even though they did not act as a deterrent. Many were positively related to announcements by local politicians concerning planned installations in their own countries (e.g., Denmark: *DR Nyheder Online*, 15 July, 2005; Russia: *Novesti*, 18 July, 2005; USA: *USA Today*, 17 July, 2005). Furthermore, both British Prime Ministers of the last decade have endorsed its benefits. In 1994, John Major stated that CCTV 'definitely' worked at crime reduction and yet this was prior to any large scale evaluation of its effectiveness (Norris & Armstrong, 1999). Furthermore, Tony Blair advocated further expansion in his New Year message of 2005 (Number 10, 31 December, 2004).

In the UK, newspapers are one of most important sources of information, due to a particularly high readership for national (68% of the population) and local or regional (84%) publications. Analyses of CCTV related stories in two national (*Daily Telegraph*: circulation 1,000,000, *The Guardian*: 387,000) and two local newspapers (*Evening Standard, Wandsworth Borough News*) for one year found the majority were positive, especially in the local newspapers (McCahill & Norris, 2002d). Articles discussing the use of local CCTV in non-motoring circumstances were generally positively presented, either in highlighting potential future crime deterrent effects, or their success in aiding the police in relation to specific offences. Few articles considered any negative connotations such as privacy or other civil rights issues. In contrast, the majority of stories in the *Daily Telegraph* (59%) were critical of CCTV, with a particular focus on a campaign against the inappropriate use of speed cameras.

Approximately 2,000,000 speed camera fines are issued per annum and the widespread implementation of further cameras was recently halted by the Government as the issue was seen as politically contentious (*The Times*, 15 July, 2005). Partly perceived by the public as primarily a means of generating revenue (POST, 2004), the Government admitted that they may be less effective than many other road safety initiatives for the reduction of accidents (*The Times*, 16 December, 2005). In addition, a recent study found that many motorists, particularly those facing a ban due to the accumulation of penalty points from a series of driving offences were avoiding prosecution by alleging that another person was actually driving the vehicle (Churchill Insurance, 14 May, 2005). In this survey, 67% of correspondents claimed that they would falsely admit to an offence in order for their partner to avoid losing their license, even though if apprehended, a custodial sentence could result. Around 700.000 drivers in a decade were believed to have evaded penalty points in this manner, which also provides some indication of public attitudes towards this type of traffic enforcement (*The Times*, 24 May, 2006).

To counteract this offence, the police in the UK recently announced that cameras will be installed that automatically take an image of a driver's face for the confirmation of identity (*The Times*, 24 May, 2006). This practice is already compulsory in order to prosecute in a number of jurisdictions in the USA. The cameras use infra-red filters to avoid dazzling drivers, particularly in the dark (*The Times*, 24 May, 2006). However, even with the highest quality images, unfamiliar face identity-matching from photographs is often unreliable (e.g., Bruce et al., 1999; Henderson et al., 2001). Therefore, unless false confessions are elicited from people of an obviously different gender, ethnicity or age to the actual driver depicted in the images, it is possible that there may be an increase in the number of cases of disputed identification in the courts. Furthermore, if more people are perceived to be evading justice, or are being wrongly prosecuted in this manner, current attitudes towards speeding cameras may become more negative.

**2. 6. Police support**

Harries (1999) suggests that for the police the primary benefit of CCTV is that of cost-saving and the ability to release images to the media. In a recent survey of 269 police officers, 90% were positive about its use, particularly in the prosecution of public order offences, theft and assault (Brandon, 2003). CCTV can induce guilty pleas, thus reducing police time as suspected criminals will often confess to a crime if they know this type of evidence exists. Indeed, in one survey it was found that all criminals informed that they were caught on camera admitted guilt immediately. Others voluntarily surrendered at police stations if local newspaper reports suggested that an image has been attained of alleged activities (Privacy International, 22 July 1997). Furthermore, researchers have stated that although they could not personally see the likeness of six convicted criminals to CCTV stills, all had pleaded guilty, when confronted with the stills during police interviews (Gill & Loveday, 2003b). However, CCTV evidence can take time to acquire and delays to court appearances are not uncommon. This can be a particular problem if a suspect claims to have an alibi, potentially able to be confirmed by analysis of video footage taken at a different site to the alleged crime (G. Davies, personal communication, 7 July 2006).

Goold (2003) also found the majority of police officers were in support of intensive CCTV coverage; particularly for confirming that they were conforming to good practice (i.e. when making an arrest). However, police officers have been suspended for suspected violent behaviour partially captured on CCTV. This has led to others being less willing to use force in circumstances that might normally require strong restraint. Goold notes that these concerns have some substance as in 1999 over 300 complaints against police officers were based on video evidence.

The introduction of CCTV can also increase the number of crimes reported, escalating police workload (e.g., Winge & Knutsson, 2003). This may be regarded as an imposition and they often prefer large retail organisations to manage their own prosecutions, particularly as the extra crimes tend to be minor, meaning resources would be directed away from serious incidents (Loveday & Gill, 2003). Indeed, retrospectively examining video footage can take many hours, meaning extensive analyses will only be conducted when investigating serious crimes. For instance, following an IRA bomb in Manchester in June 1996, police appeals for public and private films of the area prior to the attack produced over 2,000 hours of analogue footage. Norris et al. (1998) calculated that this required the inspection of 180 million individual frames. This was exceeded by the investigation into the London terrorist bombings of 7 July 2005 as the police analysed images from at least 25,000 tapes, as well as others taken using mobile phones (*Daily Telegraph*, 19 July, 2005).

**2. 7. Evaluation of CCTV schemes**

One of the problems with evaluating the effectiveness of CCTV as a crime prevention measure is with the method of assessment. With the exception of violent crime, there has been a long term trend towards reduced crime figures in the UK since peaking in the early 1990's. The British Crime Survey measured a 25% overall reduction between 1997 and 2002/3, meaning that any successes attributed to CCTV must be qualified (Simmons & Dodd, 2003). Furthermore, extra crimes are detected by CCTV that would not have previously been included in statistics and technology has also improved vehicle and building security, reducing vulnerability to crime (Winge & Knutsson, 2003).

Farrington & Painter (2003) list a number of methodological issues that should be addressed by any examination of the effectiveness of a CCTV scheme. These include the method by which crime figures are measured, ideally initially prior to announcements concerning CCTV introduction, as the publicity alone has been found to act as a deterrent (B. Brown, 1995). Evaluation of a matched control area should be conducted and they note that a reduction in crime figures may be due to other mediating variables. Many schemes are accompanied by improvements in street lighting, and changes to police and private security patrolling patterns. These can encourage the public to increase their activities in an area, acting as a further deterrent. Moreover, there is a tendency for areas with particularly high crime figures to 'regress to the mean' so that a short-term problem may be remedied in the absence of outside action. Similarly, when a specific type of crime is relatively rare in an area, small numerical changes can produce large effect sizes.

Independent studies examining the crime reduction effects of CCTV have found inconsistent evidence (B. Brown, 1995; Ditton & Short, 1998; Ditton et al., 1999; Gill et al., 2005; Griffiths, 2003; C. Phillips, 1999; Winge & Knutsson, 2003). Some have reported a reduction in property crime, but no change in the number of violent crimes (e.g., Winge & Knutsson, 2003). Others have found that the seriousness of violent or aggressive crimes is reduced, possibly due to a faster response time by the police (e.g., C. Phillips, 1999). A finding supported by the use of other more indirect measures. These include the analysis of hospital accident and emergency admissions (Sivarajasingam, Shepherd & Matthews, 2003).

The Home Office has funded a number of reviews into the effects of CCTV. The first by B. Brown (1995) comparing three city centre schemes found little impact on the type of crime that most concerns the general public (e.g., physical assault). Furthermore, a reduction in the crime statistics of two of the cities, Newcastle and King's Lynn was offset by substantial increases in Birmingham. Nevertheless, the Home Office consistently reported positive findings from this study in press releases, ignoring the negative evidence (Norris & Armstrong, 1999).

Probably the most comprehensive analysis of the effects of CCTV on local crime statistics in the UK was conducted by Gill et al. (2005). This review examined 14 different locations, sited in different environments such as hospitals, city, town and village centres and car parks. In all cases, crime statistics for at least one year prior to, and one year after implementation were analysed. There was found to be a small but significant decrease in overall crime in only two of the sites, whereas in others there was a measurable but non-significant increase, especially when breaking down crime by category. However, the authors noted that across all the schemes any effects may have been due to random fluctuations. It is apparent from these findings that if CCTV does act as a deterrent to criminal activity it is not easily detected from a relatively brief examination of crime statistics.

Indeed, the studies included in the Gill et al. (2005) review were generally short-term. None examined crime figures for more than 26 months after CCTV introduction. One study by Griffiths (2003) addressed this issue when examining the effects of CCTV in Gillingham, Kent. Crime figures in the area for 1 year prior and 5 years after implementation were compared with a CCTV-free local control town with initially similar crime rates (approximately 1,300 per annum). Reported crime fell by 44% in the first year of the scheme. However, a reduction of 22% was also found in the control area. The annual crime rate was consistently less in Gillingham throughout the assessment period and at the end of the study it remained down by 35%. In contrast, crime was only reduced by 0.05% in the control area. Based on these figures the author suggests that approximately 2,100 crimes were deterred, almost exclusively related to criminal damage and vehicle-related theft. Indeed, violent crime rose by 32% in both the surveilled area and control area, a statistic consistent with the rest of the UK in the period examined.

A related question of concern is whether the introduction of CCTV displaces crime to another location, or causes a change in the types of crimes that are reported. Indeed, Scottish police announced that CCTV coverage in towns and cities was inducing criminals to target the less secure countryside (Macaskill, 2005). The majority of early studies examining CCTV found some geographical or functional displacement (e.g., Brighton: Squires & Measor, 1996; Doncaster: Skinns, 1997;

Airdrie: Ditton & Short, 1998). In contrast, in Newcastle, CCTV installation appeared to provide a protective 'halo' around adjacent neighborhoods (B. Brown, 1995). Ditton and Short (1998) found that following the introduction of CCTV, crime was reduced by 21% in Airdrie in Scotland. Examination of these effects found that whereas some local offenders had abandoned crime, others had displaced their activities 15 miles away to Glasgow. However, perhaps more importantly, criminals from Glasgow, a much larger city, no longer appeared to target Airdrie. In comparison, a similar, but much larger-scale scheme in Glasgow was less successful (Ditton et al., 1999). The authors suggest that in Airdrie, potential criminals were generally known to operators, thus increasingly the likelihood of identification. However, in larger conurbations such as Glasgow, it is easier to remain anonymous and so CCTV is less of a deterrent.

There is also evidence that alternative initiatives may be more effective at reducing crime. For instance, Painter and Farrington (1997; 1999) found that implementation of new street lighting schemes resulted in far more positive crime reduction figures of approximately 20%, compared to the 4% reduction found by those included in the same team's meta-analysis of CCTV (Welsh & Farrington, 2002).

**2. 8. Offenders, operators and the efficacy of CCTV**

The above studies tend to support the proposal that some criminal activities may be curtailed by the introduction of CCTV. However, most of the evidence suggests that it is opportunist and less serious crime that is reduced. Gill and Loveday (2003a) questioned 77 imprisoned offenders to see if they considered CCTV to be a threat to their 'professional' activities. Those who had been filmed by CCTV cameras considered that it increased the risk of legal proceedings, although the majority perceived it to present no significant threat. They believed that coverage, film quality and monitoring are poor and that the police lacked the resources to investigate reports. They also claimed that it would often be difficult to determine if a crime was taking place, especially those involving drug dealing or credit card fraud. Indeed, in at least one notorious baby abduction case, CCTV operators wrongly believed that they were viewing a nurse acting normally (Norris & Armstrong, 1999). However, most prisoners conceded that they would have to modify their behaviour in some way, possibly by wearing a disguise or by standing with their backs to cameras.

Although offenders may believe that they can adapt their behaviour, Loveday and Gill (2003) found that experienced CCTV operators considered behaviour and body language to be the two main grounds for suspicion. Staff members trained and primarily designated as CCTV operators were extremely effective at detecting potential and actual shop theft, significantly reducing stock losses in stores. Conversely, an experimental study found that experienced control room staff and naïve participants were equally effective in predicting the occurrence of anti-social and criminal activity from recorded clips of real events (Troscianko, Holmes, Stillman et al., 2004). This suggests that the effects found by Loveday and Gill may be as a result of staff incentives and the ability to recognise repeat offenders, rather than expertise in predicting specific behaviour patterns.

### 2. 9. Civil rights and legal issues

CCTV surveillance has been discussed in terms of an updated version of Jeremy Bentham's 1787 utopian Panopticon model. This proposed that prisons, factories, workhouses or asylums should be constructed in which individuals are constantly under threat of 'permanent, exhaustive, omnipresent, surveillance', regardless of whether they actually are (e.g., Foucault, 1977). Fyfe and Bannister (1996) argue that widespread CCTV fulfills the Panopticon principle that power should be unobservable, meaning that individuals self-regulate their behaviour and deviant actions are avoided. This has the effect of ensuring conformity and facilitating the power of the observer.

Opponents of CCTV argue that this provides the state with too much power over the individual allowing it to control and penetrate civil society (e.g., S. Davies, 1999). The privacy of innocent individuals is undermined and specific groups may become closely monitored, restricting their democratic activities. These might include political or religious activists, whose campaigning or right to assembly may be viewed as acting as a deterrent to economic and retail activity. Gatherings of young people may also find that they are excluded for similar motives (Graham, 1998). However, prominent political proponents of CCTV have argued that the protection and security of the state and public should outweigh the potential loss of any civil liberties. Indeed, the Prime Minister John Major in 1998 declared 'I have no doubt we will hear some protest about a threat (from CCTV) to civil liberties. Well I have no sympathy with so called liberties of that kind" (McCahill & Norris, 2002c; p. 12).

In most European countries video surveillance is licensed and its operation is legally regulated and enforced by statute (Gras, 2004). Similarly, in the USA, privacy and the right of individuals to be protected from unreasonable intrusion is established in the Fourth Amendment to the Constitution (Maguire, 1998). In the UK, there are few legal constraints on individuals or organisations installing CCTV. No central regulatory agency has been created and no specific legislation controlling CCTV has been enacted. However, a number of separate pieces of legislation, impacting on CCTV provision have been turned into law. These are the Town and Country Planning (General Permitted Development Order 1995 Part 33, SI No. 418); the Crime and Disorder Act (1998); the Human Rights Act (1998) and the Data Protection Act (1998).

None of these laws restrained the implementation of new schemes. The Town and Country Planning order specifically permitted CCTV installation without the need for planning permission. Under the Crime and Disorder Act (1998), local government and police are obliged to coordinate crime reducing strategies and form Community Safety Partnerships. These are encouraged to apply for CCTV funding. They must only formulate a code of practice, but restrictions are limited and codes normally only specify that images should not be sold for profit or entertainment.

The Data Protection Act (1998) also legislates on the use and storage of data. To clarify the impact of the Act on CCTV, guidelines were issued so that large-scale operators register their systems (Data Protection Commissioner, 2000). As such, CCTV must only be operated for legitimate and lawful reasons (e.g., crime prevention). Signage should be in place warning of surveillance, and data should not be kept longer than necessary and should be confidential with restrictions on access. The Act also regulates on the quality of recorded images so that if the purpose of a CCTV system is to deter and prevent crime by identifying criminals then images can be collected for that purpose. However, if the registered aim is to control traffic, recording of high-quality close-up facial images may actually break the law (Taylor, 2002).

Prior to the Human Rights Act (1998) there was no constitutional right to privacy in the UK, and therefore no legal basis by which an individual could object to being filmed. This Act ensured that the UK conformed to the European Convention on Human Rights. Article 8 of the Convention specifically defined the existence of the right to privacy of individuals. As such there should be no public authority or police surveillance unless acting in the interests of national or economic security, public safety, health or morality, as well as to prevent disorder and crime and for the protection of the freedoms and rights of others (Taylor, 2002).

However, the concept of CCTV surveillance invading an individual's privacy has rarely been tested in the courts and public monitoring is not deemed to interfere with private lives (*Herbecq v Belgium*, 1998). One exception was decided in the European Court of Human Rights (*Peck v United Kingdom*, 2003). CCTV footage of the plaintiff had been released by the local authority to the national media and was wrongly portrayed as illustrating the successful use of CCTV against criminal activity. The court found that Peck's exposure in the media exceeded normal expectations and therefore interfered with his rights under Article 8. It admonished the local authority for allowing Peck to be identified and also for not seeking his consent for disclosure. To some extent therefore, this judgement supported the use of CCTV in public, and making available images of criminal activity to the media. It only criticised the release of these specific images (Gallagher, 2004).

There is also an indication of a cultural divide between the UK and other nations in attitudes towards CCTV. In contrast to the high levels in the UK, it was banned in Denmark from 1982 to 1998, with ATM (Automatic Teller Machine) surveillance only allowed in 2002, subject to strict regulations protecting the privacy of individuals (Wiecek & Saetnan, 2002). The first civic systems were installed in 1997 in the Netherlands (Flight et al., 2003), and 1999 in Norway (Winge & Knutsson, 2003). In Greece, restrictions on the use of cameras installed for the 2004 Olympics are in place, in that those designed for traffic control are not allowed to focus on faces as this is seen as an invasion of privacy (*Kathimerini*, 26 August 2005). In both Denmark (Gras, 2004) and Austria (Norris et al., 2004) the right to privacy; or the protection from state intrusion of an individual is perceived to outweigh any benefits of visual surveillance systems. Nonetheless, whereas previously restricted or banned in these countries, highly regulated schemes are being introduced and their prevalence may rapidly expand.

Norris and Armstrong (1999) argue that the rise in CCTV surveillance is in part due to the breakdown of the communist systems in Eastern Europe. During the Cold War there was a desire by politicians to characterise western society as relatively free from state interference. However, once there was no need to demonstrate this ideological difference, western authorities were able to implement intensive schemes, marketing them as a protection of the rights of individuals. Even so, experience of Nazism in Europe left the public more distrustful of state intrusion. This is especially evident in France and Denmark, as in the latter, open-space CCTV was deemed by many to be a 'spy on innocent people' (Reeve, 1998). Sutton and Wilson, (2004) list some philosophical reasons why other countries such as Australia have not embraced CCTV to the same extent. The causes of crime are perceived as having their foundation in economic and social inequality, so that if this disparity is reduced then crime rates will also fall. In this context, crime reduction schemes such as CCTV are viewed as implicitly designed to preserve the status quo, increasing the discrimination of certain groups within society.

### 2. 10. Summary

CCTV appears at present to be able to safely retain its status amongst the general public as a positive crime reduction mechanism, aiding in protecting the public and property, acting as a deterrent and complimenting other police methods for the identification of offenders. Delivery of a new system can fulfill a perceived political service to local residents and businesses by demonstrating a visible action against crime. Regardless of its effectiveness, local and national governments are required to be seen to be 'doing something' when faced by criminal or terrorist activity. Indeed, at least one local council is providing CCTV cameras to individuals targeted by anti-social behaviour and vandalism (*Peterborough Evening Telegraph*, 12 January 2005). Some observers have noted that intensive CCTV surveillance has even affected the fashion industry. Hooded garments have become a common clothing accessory, perhaps in an attempt to attain some privacy. Although few wearers are likely to be involved in criminal or anti-social acts, some private shopping malls have banned their use, a move supported by the UK Government (Steyn, 2005).

However, if communities feel that CCTV is not effective or their right to privacy is infringed; opinions may change. In which case, caution may be required before further expansion, especially in countries where coverage is currently low. Surveillance could be perceived as unnecessarily intrusive and could lead to suggestions that all members of the public are under permanent suspicion. Discriminatory monitoring has been observed. Norris and Armstrong (1999) found during an observational study of CCTV control room operators that 90% of targets were male, young and from ethnic minorities, with the homeless, vagrants and alcoholics a secondary target. Similar potentially discriminatory targeting has been found in Oslo and Copenhagen (Lomell, Saetnan & Wiecek, 2003). Nevertheless, recent UK local authority guidelines have specifically highlighted the negative connotations and in a later examination of three London control rooms in 2002, almost 50% of targets were believed to be over 30-years of age and there was no evidence of monitoring based on ethnicity (McCahill & Norris, 2003b).

However, it is possible that automatic systems in the future may contain algorithms, designed to monitor specific racial or other minority groups if they are perceived to be more of a threat. Furthermore, the proliferation of speeding cameras has been criticised as essentially a revenue generator, with doubts cast as to their contribution to road safety (POST, 2004). Others have criticised the use of traffic images in television schedules (e.g., Norris & Armstrong, 1999), ostensibly marketed as public information and safety programmes, they are perhaps perceived more as entertainment. Furthermore, if intelligent scene monitoring is utilised primarily to identify and fine minor public nuisances such as littering or vehicle infringements, further criticism may be generated. To retain the community's confidence in CCTV, cameras should "be deployed and operated with integrity and with respect to personal privacy and civil liberties" (House of Lords Select Committee on Science and Technology; 8th Report, 1997 – 1998).

It is also possible that regardless of social and political attitudes, CCTV may soon be superseded by other surveillance technologies. Mobile phones taking moving high resolution video footage were first introduced in Japan and in at least one city, Osaka; the police actively encourage the public to send them images of crime scenes (Fitzpatrick, 2002). Requests by the Metropolitan Police for mobile phone images were also highly publicised following the London terrorist bombings in July 2005 (*Daily Telegraph*, 19 July, 2005). The Forensic Science Service (2004) estimated that there were 70 million mobile phones in the UK, although only a minority had image-taking facilities. However, it is likely that the proportion with cameras will increase as companies compete for a share of this saturated market.

# Chapter 3: General Methodology

## 3. 0. Research strategy

The primary aim of the research described in this thesis was to evaluate three of the four recommendations made by Court of Appeal judges in a review of the admissibility in court of CCTV evidence for identification purposes (Attorney General's Reference, No. 2 of 2002, 2003). One recommendation was that if images were sufficiently clear, a jury could be invited to base their verdict on a resemblance to the defendant. A series of experiments were therefore implemented in which participants replicated scenarios that may be faced by individual jurors when provided with this type of evidence. In the majority, participants simultaneously matched individuals shown on video footage with either facial photographs or with live actors. The time between video capture and identification session was varied and the actors were sometimes shown on video in disguise. In a further experiment, volunteers in groups of 12 were encouraged to act the part of a jury to allow examination of deliberation processes when presented with this type of evidence.

Published research examining the simultaneous matching of unfamiliar faces has consistently demonstrated the difficulties inherent in this type of task, especially if viewpoint, expression or lighting effects are dissimilar in the two images (e.g., Bruce et al., 1999; Henderson et al., 2001). However, photographs have always been used as comparison images, which can only present a single still view of an individual. If CCTV footage depicted an incident in which a perpetrator was filmed moving, and from different viewpoints, a single photograph would not capture the multitude of detail. Therefore one objective of this series of experiments was to examine matching ability when the target was present in person. As the Court of Appeal review recommends that a contemporary photograph of the accused taken at the time of the offence should also be available to a jury, additional experiments were conducted with photographs as stimuli.

One potential criticism of using live actors as 'stimuli' in this manner is that due to the financial costs of recruitment, only a limited number can be employed and these may in some way be unrepresentative of the population as a whole. This criticism would be particularly contentious if actors had an unusual appearance, leading to a lack of external validity as findings might not reliably generalise to the wider population (Wright & Sladden, 2003). Wells and Windschitl (1999) also observed that construct validity is under threat if "a single stimulus instance from one category is used to represent one condition of an experiment and a single stimulus instance from another category is used to represent another condition of the experiment" (p. 1116).

The live actor experiments reported in this thesis could potentially be criticised in this manner, particularly for their reliance on a minimal number of actors mainly recruited from a small database of potential volunteers. Indeed, in some, only two actors actually attended identification sessions. The validity of using a small database of volunteers is expanded on in Section 3.1. However, experiments using a larger database of photographs, which included those of the actors employed for the live studies were conducted, in order to compare performance across designs. As is demonstrated throughout the thesis, the findings of both types of experiment (photographs and live actors) were generally consistent and therefore it could be assumed that the live actors employed *were* representative of the larger population and that external validity was not being violated. In addition, a pilot study, described in more detail in Section 3.7 was conducted to ensure that none of the actors taking part in any of the experiments had a particularly distinctive or unusual appearance, and therefore it was unlikely that construct validity was being violated.

In addition, the experiments in this thesis were designed, so that for instance, a comparison could be made between identification rates in one condition with those of a second condition. As is normal, statistical tests were conducted to contrast performance in those conditions and these were generally based on mean performance. However, as the main purpose of the experiments in this thesis was to test features of current legal procedures, as instructive are measures of extreme performance in individual trials.

For instance, in one experiment in the Bruce et al. (1999) series of matching studies, 20% of participants across all trials were unable to accurately match the target with their image in an array. However, in one specific trial, 80% of participants were in error. If conclusions were based on this single trial only, criticism *could* be directed at the findings for violating construct and external validity. However, in terms of the 'beyond reasonable doubt' standard of proof in the criminal law courts, this finding could be interpreted as being of more importance than that of the overall group mean of those participants. This is because it demonstrates that unless the data in that experiment was extremely skewed, some of the other trials must have had much lower error rates than 20%. These may involve targets for whom no similar-appearing distracters could be found. As such, these trials may have no relevance in a forensically-valid experiment, as they would be unlikely to represent what could occur in real life situations.

Therefore, in some of the experiments reported in this thesis, the different actors recruited as 'stimuli' are treated as a further independent variable. Although this may violate *normal* assumptions of construct validity, it allows for a statistical demonstration of the most extreme examples that could be put before the courts as evidence in forensic cases. In addition, in experiments in which performance was not compared across specific stimuli, mainly due to participants completing more than one trial, a supplementary results section was included, giving details of the most extreme error rates.

In addition, most face processing studies require participants to take part in multiple trials, providing statistically powerful data allowing measurement of subtle effects. However, in a court, a jury would rarely deliver a verdict to more than one defendant alleged to be depicted in video footage. Therefore, a further ecologically valid feature was that each participant should contribute a minimal number of data points, and in some experiments, make only a single identification decision. However, this created a further problem in that extremely large numbers of participants were required in order to ensure that each study had sufficient statistical power for the detection of even moderate effects. This issue was addressed as over 3,000 volunteer visitors were recruited at the Science Museum in London together with several hundred students and staff at Goldsmiths College, University of London.

### 3. 1. Actor selection

A feature of previous studies with a similar design was that the actors were selected from relatively small databases of individuals (Bruce et al., 1999; Henderson et al., 2001). For instance, Bruce et al. (1999) included 160 faces in their experiments from a total pool of 200 police recruits. If a larger pool had been available, error rates may have increased as it should have been easier to construct arrays with more faces likely to be mistaken for one another. Therefore, the actors selected for most of the experimenters reported in this thesis were from a social group containing fewer members than the above. The basis of this was that findings would have greater weight with a smaller database and would also avoid any criticism that a large scale 'search' had been carried out to specifically locate individuals who would be mistaken for one another.

In addition, a common finding is that members of different racial groups are worse at recognising faces of people from other racial groups (e.g., Meissner & Brigham, 2001; Valentine, 1991), especially if experience of those groups is low (Chiroro & Valentine, 1995). It was expected that most participants would be recruited from the white Caucasian majority population of the UK. Therefore, no actors from ethnic minority groups were included, so as not to deliberately enhance error rates. However, there were no restrictions on ethnic minorities contributing data as participants, as their likelihood of inclusion was expected to be commensurate with expectations of being called for jury service.

Nine paid volunteers were initially recruited to act as 'stimuli' for experiments involving live actor matching (see Chapter 6). All were members of Royal Holloway, University of London student rugby club. The following inclusion criteria were stipulated: - male, white Caucasian, with brown or black hair, neither receding nor over collar length, ears visible, a shaved appearance and with no distinguishing facial marks, unusual hairstyle or paraphernalia such as facial jewelry. They were aged between 19 years 10 months and 21 years 5 months, of slim or medium/muscular build, 72 – 95 kg, 1.70 - 1.92 metres in height; Body Mass Index (BMI) from 23.48 - 24.69, which is within the normal range (BBC Health, 7 January, 2005). At the time of filming in February 2003, from a total rugby club membership of about 60, approximately one-third met inclusion criteria.

For the experiments involving photographs, videos of the same nine actors were utilised and more were taken of a further 33 male volunteers (Chapters 4 & 5). Not all of these actors met the strict inclusion criteria. Nevertheless, all were white Caucasian, aged 18 to 25, of slim or medium/muscular build with minimal facial hair. All were either members of the same rugby club ($n$ = 15), the same college football club ($n$ = 10), or had been spectators at a rugby club match ($n$ = 8).

Four further white Caucasian male actors were recruited for Experiment 7.1, in which a different type of video image was used. These volunteers, while meeting the same inclusion criteria were all from a non-student background. They were aged 20 – 21 years, their height ranged from 1.80 – 1.85 m, weight; 73 – 80 kg.

**3. 2. Video images**

The quality of the video footage in previous studies examining matching of unfamiliar facial images has ranged from close-up head and shoulders views (e.g., Bruce et al., 1999), to images taken from a camera sited approximately 6 metres above ground level (e.g., Davies & Thasen, 2000). Home Office guidelines provide practitioners advice as to the optimum size of images in different circumstances (Aldridge, 1994). These are based on the 'Rotakin' (R) standard test target 1.6m high, intended to imitate a human (Aldridge, 1989). When the Rotakin takes up the entire vertical axis of a monitor screen, the image height is equal to 100%R. The guidelines suggest that for detection of an individual in CCTV, the target should not be less than 10%R; 50%R for accurate recognition by someone familiar with the person in the image and at least 120%R for identification by someone previously unfamiliar with an individual (assuming facial features are not obscured). For this thesis, two types of video footage were obtained, designed to meet these guidelines in different circumstances.

### 3. 2. 1. 400%R facial close-up videos

These colour videos were high-quality close-up facial portraits taken in an internal environment. The footage depicted the actors' faces slowly turning $180^0$ continuously from left-to-right profile and back replicating the 'Gold Standard' described by Bruce et al. (1999), when evaluating matching to photographs. They fulfill Home Office unfamiliar identification criteria as they measured well over 100%R (Aldridge, 1989, 1994) and they would also definitely be of 'sufficient' quality for a jury to be invited to make an identity decision without requiring further substantiating identification evidence (Attorney General's Reference, No. 2 of 2002, 2003).

Actors were filmed from a distance of approximately 2 metres using a JVC 400x Digital Zoom Compact VHS Camcorder, and were recorded on JVC Extra High-Grade VHS Compact videotape so that facial images took up three-quarters of the screen. They were asked to face the camera, turn slowly for a right profile view, followed by a left profile view and then face forward, keeping a neutral expression for about 5-sec in each view. Images were transferred using Adobe Premiere Pro Software into a digital format for playback using PowerPoint software, and edited so that they could be viewed in a continuous loop. Stills from this footage can be found in Appendix B.

### 3. 2. 2. 50%R Medium-range videos

These images were medium-range views, designed to simulate footage obtained from typical operator-controlled or automatic, externally-based CCTV open-street systems. The detail and clarity of the images used in these experiments was high, with more comprehensive views of the actors than would be found in the majority of criminal investigations.

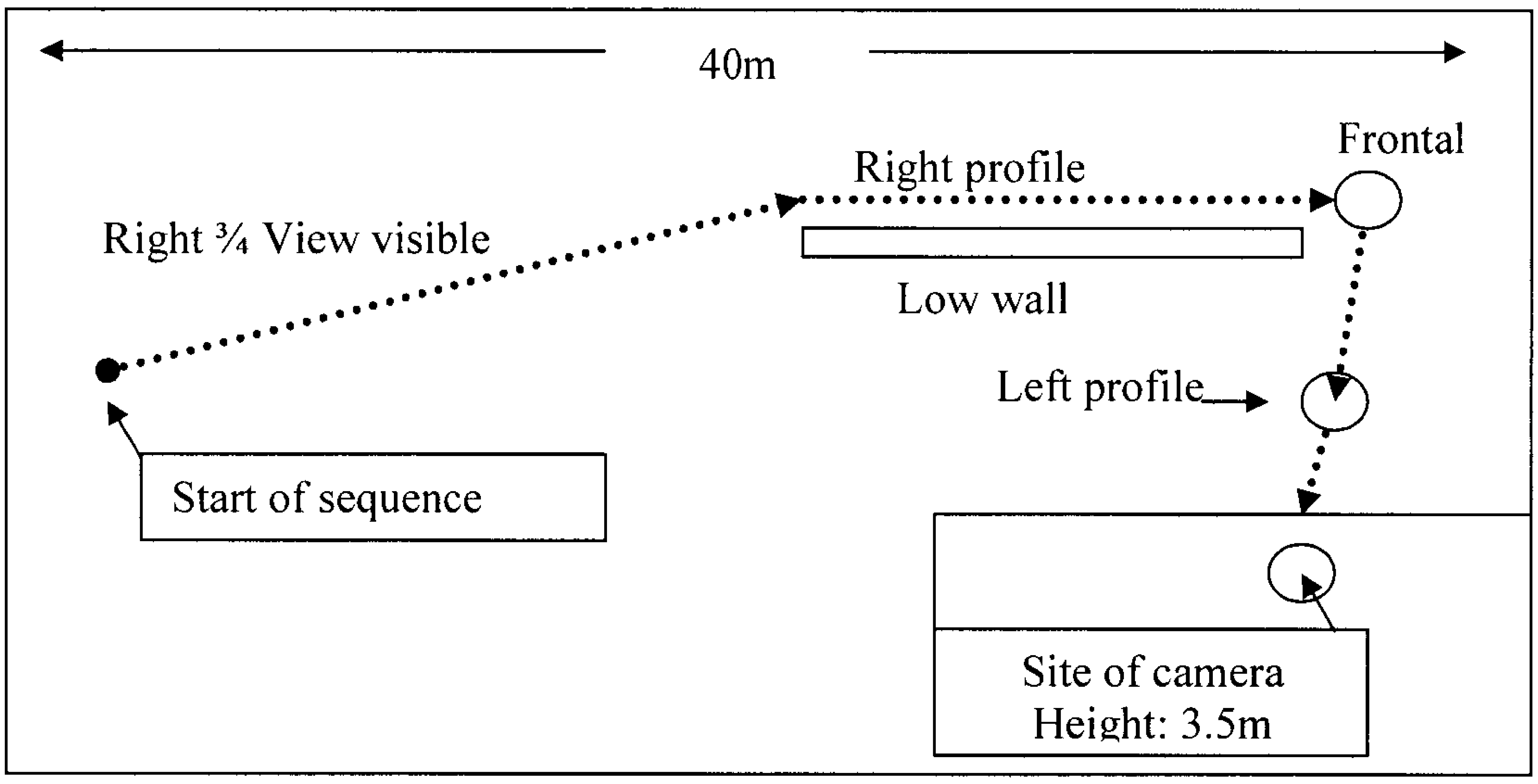



*Figure: 3. 1. Floor plan illustrating the sequence taken by the actors*

The footage was filmed so that the actors' bodies took up between one-half (50%R) and two-thirds (67%R) of the screen, allowing identity decisions to be made from more than facial features alone. These met Home Office guidelines for the minimum size for accurate recognition by an individual familiar with the target, but would not necessarily be sufficient for accurate identification by someone unfamiliar (Aldridge, 1989, 1994). The actors were individually filmed three times performing the same sequence in different conditions: in no disguise, wearing dark glasses and wearing a hat that covered the ears. The camera was a 700x zoom lens JVC VHS/C Compact Camcorder set approximately 3.5 metres from the ground, designed to simulate the operations of a manually-controlled CCTV system. The image quality was high, being that of a good home video camera. The camera was operated so that the target was maintained in medium shot by the use of the zoom and automatic focus function, while rotating the view through approximately 60 degrees.

*Figure: 3. 2. Stills of video footage taken of Actors A01 (top) and A02 (below)*

Each actor performed a 30-sec sequence of choreographed actions over a distance of 50m which included views of approximately 5-sec of the front and each side of their faces and bodies (Figure 3.0). The sequence involved walking at an angle towards the camera so that the actor's face was shown from a three-quarters view, continuing perpendicular to the camera, behind a low wall (1 metre high) so that the right facial profile was visible; standing still to the side of the wall, facing the camera, for a full-face view of the face; turning and standing so that the left profile of the face was visible; and finally walking towards and underneath the camera position. The background environmental details also provided some cues as to the height and build of the actors. Using Panasonic NV-FJ760 and JVC HR-J680 video recorders, the films were transferred and edited on to Maxell S (Super Power Tape)

VHS cassettes. Stills from the video footage of two of the actors are shown in Figure 3.1.

### 3. 3. Photographs

A set of 10 photographs, from a distance of 3 metres, with the camera on full zoom, using a Samsung Fino 1O5OXL (zoom 38 - 105mm) camera loaded with black-and-white Ilford FP4 125 film were taken of each of the 42 volunteer actors recruited for the full-body image studies (together with a further 24 participants, not videoed but included in the facial similarity study described in Section 3.7). The actors adopted a neutral expression throughout and were photographed in the three disguise conditions. This resulted in three full-face, three left three-quarters view (determined by sightline with the outer canthus of the right eye on the edge of the visible field; Appendix A) and three left-profile close-up portrait views. A single right three-quarters view facial portrait was also taken with the actors in no disguise and one further full-length colour frontal photograph was taken from a distance of 4 metres using an Olympus Trip-XB3 (lens 34mm) camera. All images were obtained within 24 hours of the videos, ensuring that hair style had not changed, although the actors wore different clothing. Photographs were scanned and transformed to BMAP files using a Hewlett-Packard HP ScanJet 3570c and Adobe Photoshop. Images were trimmed to remove extraneous cues. Image height and resolution was standardised for display purposes.

Frontal images of all 66 of the above actors can be found in Appendix E. In that database, photographs 1 – 42 depict actors whose images were used in Experiments reported in Chapters 4 – 6. Four further close-up full-face photographs taken under the same conditions and with the same camera were also obtained of the actors described in Chapter 7. These images are listed as numbers 43 – 46. The 24 frontal images of the participants who were photographed, but not videoed are denoted with numbers 47 – 70. In addition, frontal images of a further 30 actors were obtained for the facial similarity study described in Section 3.7. These images are a listed as numbers 71 - 100.

## 3. 4. Pilot study measuring familiar face recognition in the medium-range footage

Past research has consistently found high recognition rates for familiar faces in even poor-quality footage (Bruce et al., 2001; A.M. Burton et al., 1999). Therefore, a pilot study was initiated to ensure that the actors whose videos were utilised in the live full-body studies would be recognised by people familiar with them, as would be expected under Home Office guidelines (Aldridge, 1989, 1994). Eight ex-members of the same rugby club were recruited. None were previously aware of the involvement of the actors who had volunteered to be filmed. Each was shown the video footage of the nine actors wearing a hat. In all cases, within a few seconds, naming accuracy was 100%.

## 3. 5. Pilot study to select distracters for live actor experiments

For experiments involving the nine live actors (Chapter 6), participants were recruited for a second pilot study from an amateur dramatics society ($n$ = 20). All were unfamiliar with the target actors and none took part in further experiments. They paired each actor with one other based on physical resemblance using the full body colour photograph taken at the time of filming. They were required to select the pair they believed would be most likely to be mistaken for each other under poor viewing conditions, and subsequently pair the remainder. The distracter for each live actor in target absent conditions was the video actor that had been selectively paired by the most pilot participants. Using this method, one distracter actor was chosen as being highly similar to two different target actors. Therefore, the video of this distracter was played alongside these two live actors in target absent sessions (see Section 6.1.1.2).

## 3. 6. Pilot study to select distracters for photographed actor experiments

A third pilot study was conducted in which 40 undergraduate students from Goldsmiths College, University of London produced matched-pairs and arrays for experiments that involved identification from photographs (Chapters 4 & 5). Twenty of these participants were given full-face views, the remainder right three-quarters facial images. Following a procedure described by Bruce et al. (1999), participants were asked to organise the 42 photographs into piles based on perceived similarity. There were no limits as to the number in each pile and any actor appearing 'unique' could be set aside. The number of piles produced by participants ranged from 4 to 14, with a median of 9 and a mode of 10. This data was entered into a 42 x 42 matrix showing the frequency with which any one face was paired with another. For experiments in which participants were to compare a single photograph with a video image, 12 pairs of faces were selected that had the highest frequencies in the matrix. No face was paired with another more than once. In three cases, constructed pairs were the same as those for the live actor matching experiments.

### 3. 7. Pilot study rating actor photographs for distinctiveness and similarity

Many recognition and eyewitness studies have found that people rated as distinctive in appearance are more likely to be correctly identified if present in a line-up/array, and are also less likely to be falsely identified if the correct target is not present (e.g., Light et al., 1979; Valentine, 1991). The deliberately homogeneous criteria of the inclusion policy for the actors being filmed in the studies in this thesis meant that it was unlikely that any would 'stand out in a crowd'. However, within this group some may have been more distinctive than others, potentially impacting on performance. Seventy-five further participants from Goldsmiths College, University of London were asked to sort 65 out of 100 facial photographs into a similarity matrix using a similar method as described in Section 3.6[1]. These photographs included the 46 full-face photographs described in Section 3.3, as well as a further 54 photographs obtained of males meeting the inclusion criteria, and are all depicted in Appendix E. This data was entered into a 100 x 100 matrix showing the frequency with which each face was paired with another. From this procedure, two faces were paired 25 times, the highest frequency in the matrix. The face with the

---

[1] To sort 100 faces was found to be too onerous a task for participants. Therefore, 65 faces were randomly chosen for each, with the proviso that the 35 not given to a particular participant would always be provided to the next.

highest number of matches was paired with 61 others; the lowest was paired with 16 others. The full results of this procedure are discussed in the relevant chapters.

### 3. 8. Single-item identity-verification design

In the majority of experiments, a single-item identity-verification design was employed. This replicates the scenario whereby a juror would be required to decide if a person shown on video was either physically present or alternatively depicted in a photograph. Target present and target absent trials were conducted (Figure 3.3).

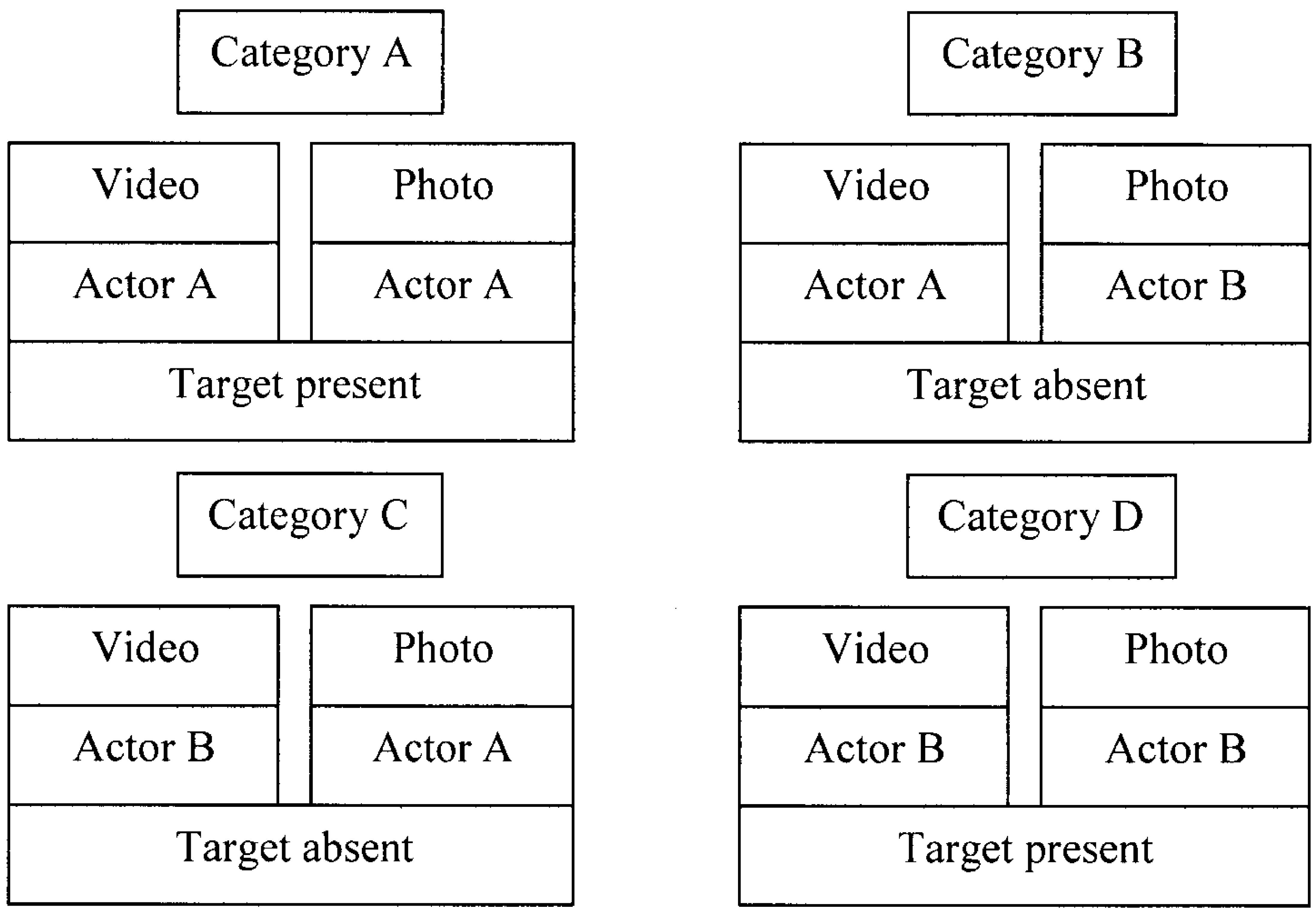



*Figure: 3. 3: Category structure for target present and target absent conditions using the same two matched actors*

The method of construction was such that if a particular actor was present in the first video, the actor matched highest for similarity of appearance by pilot participants would appear on the second. In most experiments, participants completed six trials and for these, three of the six actors present in each film were also shown in photographs for target present conditions. For the other three target absent trials, a photograph of a distracter was present instead. For instance, (a) Actor A could be present in the video and photograph (target present); (b) present in the video, with Actor B (the distracter rated highest in similarity) in the photograph (target absent); (c) present in the photograph, with Actor B in the video (target absent); or (d) Actor B could be present in both video and photograph (target present). Participants would be presented with one of these four different categories only. However, categories were always fully rotated across different participants.

Within the eyewitness study literature, there has been a debate as to whether data from target present and target absent trials should be combined in a single analysis (e.g. Lindsay et al., 1997). Responses in both categories can be classified as correct vs. incorrect. However, a correct identification in a target present trial requires selection of the correct target, whereas a correct rejection in a target absent trial relates to a failure to choose any face. This may involve different psychological mechanisms. Most commentators tend to argue therefore that target present and target absent trials should be analysed separately. This is probably a correct policy in memory experiments involving line-ups, as the number of distracters in each line-up will have confounding effects as will the resemblance of any particular distracter to the target.

However, in the majority of experiments reported within this thesis, in which memory was not assessed, the target present and target absent conditions were fully counterbalanced so that the same actors and their matched distracters were included in both types of trial. Therefore, all data was analysed together, as often an examination of the main effects and interactions involving target presence as a factor were found to be instructive. However, simple effects analyses were also conducted, treating the conditions as two separate variables.

### 3. 9. Eight-point identity-decision and confidence scale and statistical analyses

In eyewitness research the correlation between recognition accuracy and confidence has often been found to be relatively weak (e.g., Bothwell, Deffenbacher & Brigham, 1986; Sporer et al., 1995). This may partly be the result of experimenters ensuring that to the best degree possible there is little variability in procedures across participants, meaning that accuracy rates are fairly similar. The findings of eyewitness studies do not directly impact on those examining face matching. However, in cases of disputed identity, it would be likely that individual jurors with high confidence in their ability to decide if a perpetrator on video is the defendant would vote accordingly.

| **Age:** | | **Gender:** | | |
|---|---|---|---|---|

Please circle one of the numbers indicating the extent that you believe the person seen in the video is the person you see in the room. Please do not discuss this with anyone else.

| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
|---|---|---|---|---|---|---|---|
| ↑ Definitely **NOT** the same person | | | | | | | ↑ Definitely the same person |

*Figure: 3. 4: Eight-point identity-decision scale*

Therefore, in all the single-item identity-verification design studies reported in this thesis (Figure 3.4), participants were asked to record their responses on an 8-point identity-decision scale ranging from 1 (definitely not the same person) to 8 (definitely the same person). The scale allowed for categorization of decisions as either 'same' (5, 6, 7 and 8) or 'different' (1, 2, 3 and 4), meaning accuracy could be assessed, in which case incorrect responses were given the value of 1, correct the value of 0. Scale scores were also recoded so that those on the extremes (e.g., 1 & 8) were taken to indicate a high confidence in 'different' or 'same' decisions respectively, and coded 4 on a confidence scale of 1 – 4. Those near the centre (e.g., 4 & 5) were treated as unsure decisions and recoded as 1 on this scale.

However, even though large numbers of participants were recruited in most experiments, a dichotomous dependent variable based on the converted scale accuracy data will inherently have low variability and thus low statistical power.

This is of particular concern when participants made a single identity decision, contributing one data point to the analyses. Provisional data screening also found that data tended to violate assumptions of parametric tests, possessing heterogeneity of variance, and varying from a normal distribution. The unconverted 8-point scale data provided more sensitivity, although due to its essential bi-modal nature (participants tended to respond more often at the extremes of the scale than at the centre) it also violated the same assumptions.

The accuracy data was therefore preferred as it could easily be converted to percentages to provide clear descriptive statistics and would provide compelling evidence if required in a court room setting to instruct jurors if a specific case called for an expert witness to present data. However, if the unconverted scale responses provided additional information, these analyses were also reported. Throughout the thesis, even though the assumptions of parametric tests were often violated, the results of these tests were fully analysed. In *all* cases, comprehensive non-parametric equivalent tests were performed on the same data, and in *all* cases the results closely corresponded with the parametric results. Therefore, for brevity, only the results of the parametric tests are reported.

Due to the applied nature of the research, analyses of statistical power are not reported in this thesis. As previously noted, with a dichotomous dependent variable and with each participant contributing a single or a minimal number of data points the power will always tend to be low. In many of the experiments large numbers of participants were recruited to offset this disadvantage. With such large numbers of participants, the reporting of percentage error rates is of more value than any power indices. Indeed, the philosophy behind the enactment of criminal law in the UK and in many other countries is that a guilty verdict in court should only be rendered if it is proved 'beyond reasonable doubt' and not from any statistical balance of probabilities. Therefore, if there are any reasonable questions as to the reliability of the procedures involved in an attempt to establish guilt, there should be an acquittal verdict. Furthermore, if evidence were to be provided in court as to the likelihood that a jury could be mistaken when attempting to match the defendant with video footage, to ensure understanding by a jury, it is likely that explanations would be based on percentages and not on more complex mathematical computations.

In addition, signal detection theory based indices could also be calculated, from conversion of the scale data, dependent on the number of data points contributed by participants. To measure the sensitivity of responses when participants contributed data to *more* than one condition, the non-parametric signal detection theory measure $A^{/}$ was calculated. This statistic is the equivalent of the parametric measure $d^{/}$ in combining hit rates and false alarms using a graphical method to approximate the area under the Receiver Operator Characteristics curve (ROC). Rhodes (1993) argues that $A^{/}$ is superior to $d^{/}$ as a statistic, as being non-parametric, no assumptions about the normality of distributions are required. Values of $A^{/}$ range from 0.50, which denotes a diagonal line on the ROC curve corresponding to chance performance to scores of 1.0 indicating perfect accuracy. To control for hit rates and false alarms of 1 and 0, values were adjusted to .999 and .001 respectively. $A^{/}$ therefore provides a criterion-free measure of sensitivity in discriminating between faces (Snodgrass & Corwin, 1988) and in the identity-verification experiments reported in this thesis it acts as a measure of the ability of participants to identify whether the same target was shown in both images.

To measure bias in responses, when participants contributed data to multiple conditions, two procedures were utilised. The use of the unconverted scale data does provide this type of information, in that a high score on the eight-point identity-belief scale is indicative of a liberal response (when compared to responses in other conditions), a low score indicative of a conservative decision criterion. However, the complimentary orthogonal non-parametric measure to $A^{/}$ was also measured ($B^{//}$; Snodgrass & Corwin, 1988). In these experiments, the $B^{//}$ index is a measure of the strictness of the criterion used by participants to determine the presence of the same person in two images. As such it measures their reluctance to guess and respond 'same' to faces that look similar based on this criterion. $B^{//}$ varies between +1 and -1 with values of 0 indicating a neutral criterion. Negative values denote a liberal criterion, or a bias to respond 'same'. Positive values denote a conservative criterion, or a bias to respond 'different'.

It is possible to measure sensitivity and bias when participants contribute a *single* data point as occurred in the experiments described in Chapters 6 and 7, by using the accuracy and confidence data derived from different participants to measure the area under the ROC curve. However, as participants do not contribute data *across* different conditions, it is not possible to evaluate the specific criterion or weight the same individual might place on the different types of trial. It is therefore not as informative or powerful, and indeed for measuring *bias*, the unconverted scale data is as instructive, therefore only this is reported.[2]. For measuring *sensitivity* in the single data-point experiments, the accuracy and confidence level data produced from conversion of the identity-decision scale scores were utilised. Thus, correct and incorrect responses were separated and analyses were conducted on the confidence across each experimental condition as a function of accuracy.

Finally, an alpha level of 0.05 was set in all statistical analyses across all experiments as to the likelihood of committing a Type-I error. However, in the course of the thesis, test results well within this value, being highly significant are acknowledged. Furthermore, due to the inherent low statistical power in the dependent variables, marginal effects below an alpha value of 0.1 are reported and treated as indicative of a non-significant trend, especially if supported by similar results using complimentary analyses or across different experiments.

---

[2] As is reported later in Experiments 4.2, 5.1 and 5.2, the results from the bias $B''$ measure corresponded closely with those using the unconverted scale scores, confirming the safety of using the scale scores only when measuring bias in single trial experiments.

## Chapter 4: The matching of unfamiliar faces by adults and children

### 4. 0. Introduction

Following acquisition of CCTV footage of an incident, police officers or other security officials may use it in an attempt to identify the perpetrator, perhaps by comparing them to 'mug shot' photographs. Once a suspect has been apprehended and charged, a jury in court may also be invited to decide whether the defendant is the person shown in the footage, in which case a contemporary photograph taken at approximately the time of the crime should also ideally be available for the jury to examine (Attorney General's Reference, No. 2 of 2002, 2003).

Two experiments by Henderson et al. (2001) confirm the difficulties inherent in this task. In Experiment 1, participants were required to match CCTV stills of two different target 'culprits', whose faces were simultaneously presented within arrays of 8 high-quality frontal facial photographs. The image quality of the stills varied but was considered to be "typical of most high street banks" (p. 447). Participants were deliberately misinformed that the targets might be absent from the arrays. However, if, on their first choice they incorrectly responded 'not present', a second opportunity was given. Only 20% of participants accurately identified the correct targets with their first choice, improving slightly to 28.5% with a second choice. The remainder selected distracters.

A follow-up experiment (Experiment 2) was conducted to ensure that the high error rates were not due to a lack of detail in the stills, by replacing them with high-quality frontal photographs, so that target culprits were depicted with "a slightly different hairstyle and facial expression taken under different lighting conditions" (p. 452). Only 16% of participants correctly selected one of the culprits on their first attempt. Indeed, 40% selected a distracter, with one foil in particular chosen as many times as the correct target. When the 'not present' response was disallowed for second attempts, more than two-thirds still made incorrect selections. Similar results have been obtained with high-quality images in other experiments using the same type of design, in which participants simultaneously matched ten faces in arrays with single video images (e.g. Bruce et al., 1999).

The first experiment for this present thesis was a modified replication of those reported above. All participants, some being under the age of 18, were unfamiliar with the actors in the images and performance was compared across age groups. Older people tend to be worse than young adults at recognising faces (Pike, Brace & Kynan, 2002; Searcy et al., 1999; Smith & Winograd, 1978; Searcy, Bartlett & Memon, 1999; Valentine et al., 2003), being more prone to making false identifications when targets are absent, most notably once over the age of 50 (O'Rourke et al., 1989; Pike et al., 2002; Smith & Winograd, 1978). This may partly be due to a general cognitive decline, but seems to also be the result of the operation of an own-age face advantage effect, found in both children and adults (e.g., Chance, Goldstein & Anderson, 1986; George & Hole, 1995; Smith & Winograd, 1978; Wright & Stroud, 2002). For instance, Wright and Stroud (2002) found that actors in their early 20's were less likely to be recognised by 40 – 55 year-olds than by participants of approximately the same age as the targets (18 – 33). In contrast, the older adults were better at recognising targets aged approximately 50.

Jurors up to the age of 70 are randomly selected from the electoral roll in the UK. However, for a variety of reasons, including greater population mobility, proportionally fewer younger adults appear on this register (Henn & Weinstein, 2002). The findings of age-based deficits in face recognition are of concern as the majority of crimes are committed by younger adults. No published studies appear to have compared different adult age groups on a simultaneous face matching task. However, if similar results were found to those in recognition studies it would suggest that a jury predominantly made up of older people would be more likely to believe that a young defendant is shown in CCTV footage.

Furthermore, no studies appear to have directly compared adults and children at face matching. Although children will not be called for jury service, their performance at a task of this nature relates to their ability to recognise an offender after witnessing a crime. Studies using eyewitness paradigms have found that when targets are present in line-ups, adult levels of performance have been reached by 6 years-of-age (e.g., Goodman & Read, 1986; J.F. Parker et al., 1986). However, children tend to make more selections than adults, increasing the number of target absent errors, a strategy that conversely reduces the number of target present errors (e.g., Lindsay et al., 1997; J.F. Parker & Carranza, 1989; J.F. Parker & Ryan, 1993). Face matching performance also appears to develop throughout early childhood. Ellis (1992) found improvements from 3 - 8-years of age and was close to ceiling by 11-years of age. Furthermore, Bruce et al. (2000) examining children from 4- to 10-years of age on a two-alternative forced-choice matching task, found that face discrimination skills did not reach ceiling in 10-year-olds, the oldest tested. Nevertheless, Davies (1996b) identified specific examples of child witnesses, as young as 3 years-of-age, who have been able to select previously unfamiliar criminal offenders from deliberately-challenging photographic arrays. These often extremely serious case studies suggest that in certain circumstances, findings from face recognition studies have no bearing on the likelihood of a specific child of any age to positively identify the target.

However, Davies (1996b) summarises four general explanations for the improvement in children's performance often found at face recognition tasks. Firstly, information processing accounts suggest that encoding and storage strategies become more efficient as a child gains experience in increasing numbers of faces. Within this approach, the explanation for an increasing ability to *match* faces is that children develop the aptitude for discriminating between different faces by isolating specific features in a novel face. A second associated explanation is that children develop an expertise with faces so that an increased knowledge of what constitutes an unfamiliar face will help them in this type of task. Davies notes a number of studies finding that children perform better than adults on specific recognition tasks if they have more knowledge of, or interest in, the objects being remembered. An increasing self-awareness of what is required in a specific face recognition task has also been cited as an explanation for improved performance. As such, the development of meta-cognitive skills should allow older children to understand the likelihood of a target to actually be displayed within an array in both recognition and matching tasks. Finally, the interpretation of the social and contextual factors in a task will depend to some extent on the likely levels of sophistication required. Therefore, younger children in an unfamiliar context or setting, or when exposed to an unfamiliar experimenter may be adversely affected and for instance be more likely to misinterpret the experimental demands. Thus, the explanation for an increased number of target absent distracter selections in younger children is that the suggestiveness of the question requires a positive, even if incorrect response. As such, the interrogator in a real case should ensure that questioning is devised in a manner that reduces the likelihood of a child misconstruing the task demands.

However, there is also evidence of a leveling of performance; or even a regression in face processing skills at the start of adolescence (Carey et al., 1980; Diamond et al., 1983; Flin, 1980, 1985a; Soppe, 1986). Flin (1980) tested face recognition in children aged from 6- to 15-years. Accuracy improved from 6- to 10-years of age, with a moderate reduction in performance among the 11- and 12-year-olds. Improvements then continued from the age of 13. Soppe (1986) replicating these findings, found that other tasks such as intelligence tests and letter, figure, and handwriting recognition all *continually* improved across the same age range. Evidence for a performance dip in face matching was also found by Carey et al. (1980; Experiment 3) using the Benton and van Allen (1968) neuropsychological diagnostic test. In this task, participants are shown a series of target facial photographs above an array of six different photographs. The same target face is shown over three trials: firstly once in the array, secondly, thrice from different viewpoints, and finally, thrice under different lighting conditions. The aim is for participants to identify all seven targets. The authors combined their data with that from the original test publication, which examined children aged 6 to 11 as well as adults. Continual improvement was found up to the age of 10, followed by a plateau until the age of 14, with 16-year-olds performing at adult levels. These results imply a longer dip in matching performance than that found in recognition tasks.

Hormonal influence causing physiological changes in the brain at puberty has been suggested as a potential cause for this hiatus in face processing skills (Carey et al., 1980; Diamond et al., 1983; Soppe, 1986). Diamond et al. (1983) found that girls in the active stages of pubertal development were worse at face recognition than those matched for IQ and age but who were either pre- or post-pubertal. In comparison, performance on another visuo-spatial task was not dependent on pubertal status. Soppe (1986) suggests that early adolescents are more aware of their own and others physiognomic 'erotic' development, inducing attentional mechanisms to focus on facial features not directly of use in recognition. However, boys tend to reach puberty later and no studies have found an interaction between the performance of girls and boys at different ages as would be expected by this model (Chung & Thomson, 1995). Furthermore, the performance dip is not confined to face recognition. Temporary reverses in ability are found in voice (Mann, Diamond & Carey, 1979), flag and picture recognition (Flin, 1985b).

An alternative information processing explanation suggests that once a certain proficiency in face processing is reached, there is a subsequent internal reorganization to a more effective cognitive strategy. This takes time, causing a temporary reduction in performance (Carey et al. 1980; Chung & Thomson, 1995; Flin, 1985a). Carey et al. (1980) propose that this strategy change might be a result of encountering more faces following progression from small junior to larger senior schools at this age. However, this model cannot explain the specific relationship between puberty and performance found by Diamond et al. (1983) as the children were recruited from the same schools, and would have experienced an institutional change at the same time.

**Experiment: 4. 1**

The first experiment in this thesis was designed to directly compare adults of different ages, adolescents and children in a face matching task using the medium-range full body video images described in Chapter 3. To avoid the necessity of producing special instructions for young children, the minimum age tested was 8-years. Participants were required to match each video image with six facial photographs presented in an array. On two of the six trials the target was absent from the array. There were a number of hypotheses based on previous research:

The first was that there would be a high proportion of errors in this task as found in previous studies examining simultaneous face matching from arrays (e.g., Bruce et al., 1999; Henderson et al., 2001).

Secondly, due to both evidence of a negative correlation in adults between age and accuracy on a number of face identification tasks (Lindsay et al., 1997; J.F. Parker & Carranza, 1989; J.F. Parker & Ryan, 1993; Smith & Winograd, 1978), and to the specific effects of the own-age face processing advantage (e.g., Chance et al., 1986; George & Hole, 1995; Smith & Winograd, 1978; Wright & Stroud, 2002), older adults and children were expected to perform worse than younger adults, as the target actors were all aged between 18 and 25.

A final hypothesis was that a performance deficit or plateau, found in previous face processing studies (e.g., Carey et al., 1980; Diamond et al., 1983; Soppe, 1986) would also be observed in children at the onset of adolescence.

### 4. 1. 1. Method

#### 4. 1. 1. 1. Participants

Four hundred and twenty (164 male, 256 female) participants with normal vision took part in this study. One hundred and seventy-eight were under the age of 18 ($M$ = 11.0, $SD$ = 2.2), the remainder were 18 and above ($M$ = 36.1, $SD$ = 12.1). All were visitors to the Live Science exhibit within the Who am I? Gallery; Science Museum; London. Adult participants signed a consent form giving brief experimental details. Informed consent from a parent or legal guardian was required for those under the age of 18. None of the participants had taken part in any of the pilot studies.

**4. 1. 1. 2. Materials**

The photographic stimuli were the high-quality three-quarter view black-and-white portraits of 42 actors wearing no disguise described in Section 3.3. These were arranged into six separate sub-groups of seven photographs. Each sub-group was constructed so that to the best possible degree, actors rated as most similar in appearance by pilot participants were placed together (see Section 3.6), with no photograph appearing in more than one sub-group. Photographic arrays were constructed from six of the seven photographs from each of these sub-groups (the photographs used in each sub-group are depicted in Appendix A).

The video clips of 24 of the 42 actors wearing no disguise described in Section 3.2.2 were employed. These were transferred onto four separate VHS video tapes, so that six actors (one from each of the six sub-groups) were shown on each tape. In their trials, each participant was required to view one of these four tapes only. This meant that videos of four of the seven photographed actors in each sub-group were employed. These four actors had received the highest similarity ratings within that sub-group by the pilot participants.

In target present arrays the photograph of the target actor shown on video appeared among the five other distracter faces from that specific sub-group. For target absent arrays this photograph was replaced with a further distracter possessing the lowest similarity rating within that sub-group (i.e., the seventh photograph in that sub-group). Figure 4.1 provides an illustration of the conditions experienced by participants in target present and target absent conditions.

In summary, from any specific sub-group of seven actors, one of four could be a target and therefore could be depicted on video in either target present or target absent conditions. If it was a target present trial, the actor in the video would be depicted in the array. The other three potential targets would *always* appear as distracters, as would two further actors who were never depicted as targets on video. These actors were rated as fifth and sixth most similar in appearance ratings within that sub-group. If the trial was target absent, the actor on video would be replaced by the distracter rated seventh in that sub-group. The full list of target actors and distracters within each sub-group is provided in Appendix A.

Finally, to ensure that no bias or confounding effects were introduced by the use of four separate video tapes, each depicting different sets of actors, these were treated as a further variable in all analyses, described as the *video presentation block* condition.

#### 4. 1. 1. 3. Design

This experiment utilised a face verification design in which participants were required to match single adult male target actors shown in a video with arrays of six facial photographs. Each participant took part in six trials. Four of the trials were target present. In the remaining two trials the target was absent from the arrays and was replaced by a further distracter. The independent variable was the age group of the participants. The dependent variables were the different identification outcomes. In target present conditions these were recorded as either the number of correct hits, incorrect distracter selections, or misses (a wrongful belief that the target was absent from the array). For target absent trials, the dependent variable was the number of correct rejections.

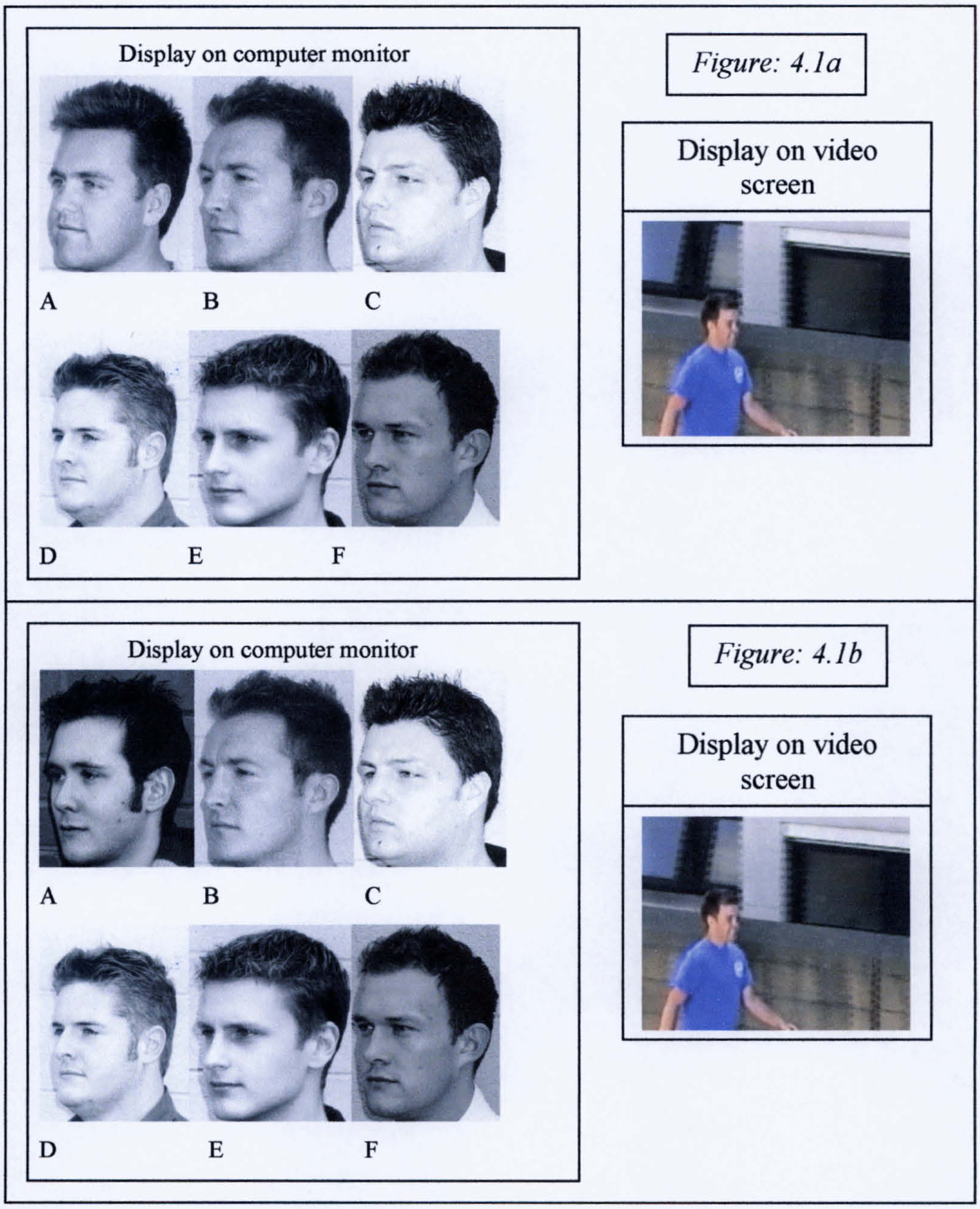



*Figure 4.1: Demonstration of conditions experienced by participants in Experiment 4.1 with top (a) depicting a typical target present trial in which Actor 36 shown in the video is within the array (position A). In the lower display (b), Actor 36 has been replaced by Actor 42 for target absent conditions.*

#### 4. 1. 1. 4. Procedure

Participants were tested in groups of up to eight, seated in a semicircle approximately 1.5 metres from a central 16" colour television/video player. Two computer monitors (20" and 22"), connected so that the same photographic image was shown on both, were sited on either side of the television. Participants viewed the series of six video films on the television screen and were simultaneously presented with arrays of photographs on the monitors. A verbal warning was given in advance, informing participants that, 'on at least one of the trials, and maybe more, the actor shown in the video will not be displayed in the array'. They were required to select from each array the photograph they believed matched the video actor, or to indicate that the target was absent. A response form was provided to record decisions. There was no time limit or constraint on the number of times videos could be viewed. However, in the majority of trials participants watched each video three times.

Arrays were presented using PowerPoint displays so that each photograph took up approximately one-sixth of the monitor screen. Participants viewed one of the four videos only. Presentation order was randomised, and counterbalanced so that the four target actors depicted in each video block were shown an equal number of times across participants. Items were therefore completely rotated and counterbalanced across all conditions. Participants were asked to be silent and were observed closely to ensure that none were discretely looking at the responses of others or collaborating in any manner. Full performance feedback was provided following the final trial.

### 4. 1. 2. Results

For an initial examination of the effects of age on responses, participants were sub-divided into four groups. These were based on two separate median splits, one of those aged 18 and above, the second of those below the age of 18. A disparate proportion of 11 year-olds ($n$ = 38) meant that it was not possible to ensure equal numbers in each age group. However, a chi-squared test found that the numbers from each age group assigned to each of the four video blocks did not significantly differ, *Pearson* $\chi^2(9, n = 420) = 16.04, p > .05$. The proportion of correct and incorrect responses (expressed as percentages with standard deviations) for target present trials and correct rejections for target absent trials for each age group are presented in Table 4.1. Of the 420 participants, 17 made the maximum number of six errors. Only four participants were 100% correct across all six trials.

#### 4. 1. 2. 1. Comparison of overall age groups

The mean hit rate for the different video presentation blocks in target present trails ranged from 43.1% in one, to 58.6% in another, indicating that some were easier than others. Therefore, this factor along with gender as a second factor was initially entered in all the analyses reported below. In all cases, whereas the main effects involving video block were significant ($p < .05$), neither the main effect of gender, nor any of the interactions involving gender or video presentation block with age group were significant ($p > .1$). Therefore responses across conditions were collapsed. In four of the six trials the target was present. Inspection of these results found that whereas 42.6% of the 18 – 38 year-olds correctly identified at least three out of four targets, only 21.7% of the 8 – 10 year-olds and approximately 30% of the two other age groups were this successful. Indeed, 41.9% of the 8 – 10 year-olds correctly identified one target or less, with 20.5% of the 18 – 38 year-olds and 32.5% of the other age groups performing at this level.

A series of independent-measures ANOVAs were conducted on the data from Table 4.1 to examine the effects of age group on each type of target present response, and on target absent correct rejections. Post-hoc analyses were performed using the Games-Howell test, recommended for when cell numbers are unequal (Howell, 2002). The first one-way ANOVA found a significant effect of age group on the proportion of correct hits when targets were present, $F(3, 416) = 5.19, p < .005$. Post-hoc tests showed that this was due to greater accuracy by the 18 – 38 year-olds than both the 8 – 10 year-olds ($p < .01$) and the 11 – 17 year-olds ($p < .05$). There were no differences between the two children's age groups. However, the younger adults also selected marginally more correct targets than the older adults ($p$ = .062).

*Table 4.1: Percentage of correct and incorrect responses in target present conditions; and percentage of correct rejections in target absent trials as a function of age group*

| | | Target present | | | | | | Target absent | |
|---|---|---|---|---|---|---|---|---|---|
| Age group | *n* | Hits | *SD* | Incorrect | *SD* | Miss | *SD* | Correct rejections | *SD* |
| 8 – 10 | *74* | 43.9 | *24.4* | 42.6 | *25.4* | 13.5 | *17.2* | 34.5 | *32.0* |
| 11 – 17 | *104* | 47.1 | *26.1* | 33.9 | *22.6* | 19.0 | *18.6* | 35.1 | *32.6* |
| 18 – 38 | *122* | 56.8 | *24.5* | 27.0 | *21.0* | 16.2 | *20.1* | 36.5 | *34.0* |
| 39+ | *120* | 49.2 | *22.7* | 33.5 | *23.0* | 17.3 | *17.4* | 36.2 | *33.7* |
| Overall | | 49.9 | *24.8* | 33.3 | *23.3* | 16.7 | *18.5* | 35.7 | *33.1* |

The second ANOVA found a significant effect of age group on incorrect selections, $F(3, 416) = 7.17, p < .001$. Post-hoc analyses found that significantly more errors of this type were made by the 8 – 10 year-olds than by the young adults ($p$ < .001). The 8 – 10 year-olds also performed marginally worse than the 11 – 17 year-olds ($p$ = .092), and the older adults ($p$ = .066). In addition, the 11 – 17 year-olds selected marginally more incorrect distracters than the young adults ($p$ = .091).

However, there were no effects of age group on the number of target present misses, $F$ (3, 416) = 1.34, $p$ > .1 or target absent correct rejections $F$ < 1. These final results illustrate that there were no effects of age group on the number of times participants selected the 'not present' option on the response sheet (combined misses and correct target absent rejections).

#### 4. 1. 2. 2. Subdivision of the age groups

To further examine the effects of age group on this matching task the participants under the age of 18 were sub-divided into three approximately equal age groups (8 – 9 years-of-age, $n$ = 53; 10 – 11, $n$ = 59; 12 – 14, $n$ = 53). To prevent confounding the results by incorporating the comparatively low number of 15 – 17 year-olds ($n$ =

13), these were excluded from this analysis. Two adult groups were also formed. One between the ages of 18 and 23 acted as a control ($n$ = 57), as this group was formed of participants approximately the same age as the actors. The second adult group was comprised of all participants above the age of 45 ($n$ = 53). Figure 4.1 illustrates the percentage of hits, misses and incorrect selections in target present conditions for each of these sub-divided age groups.

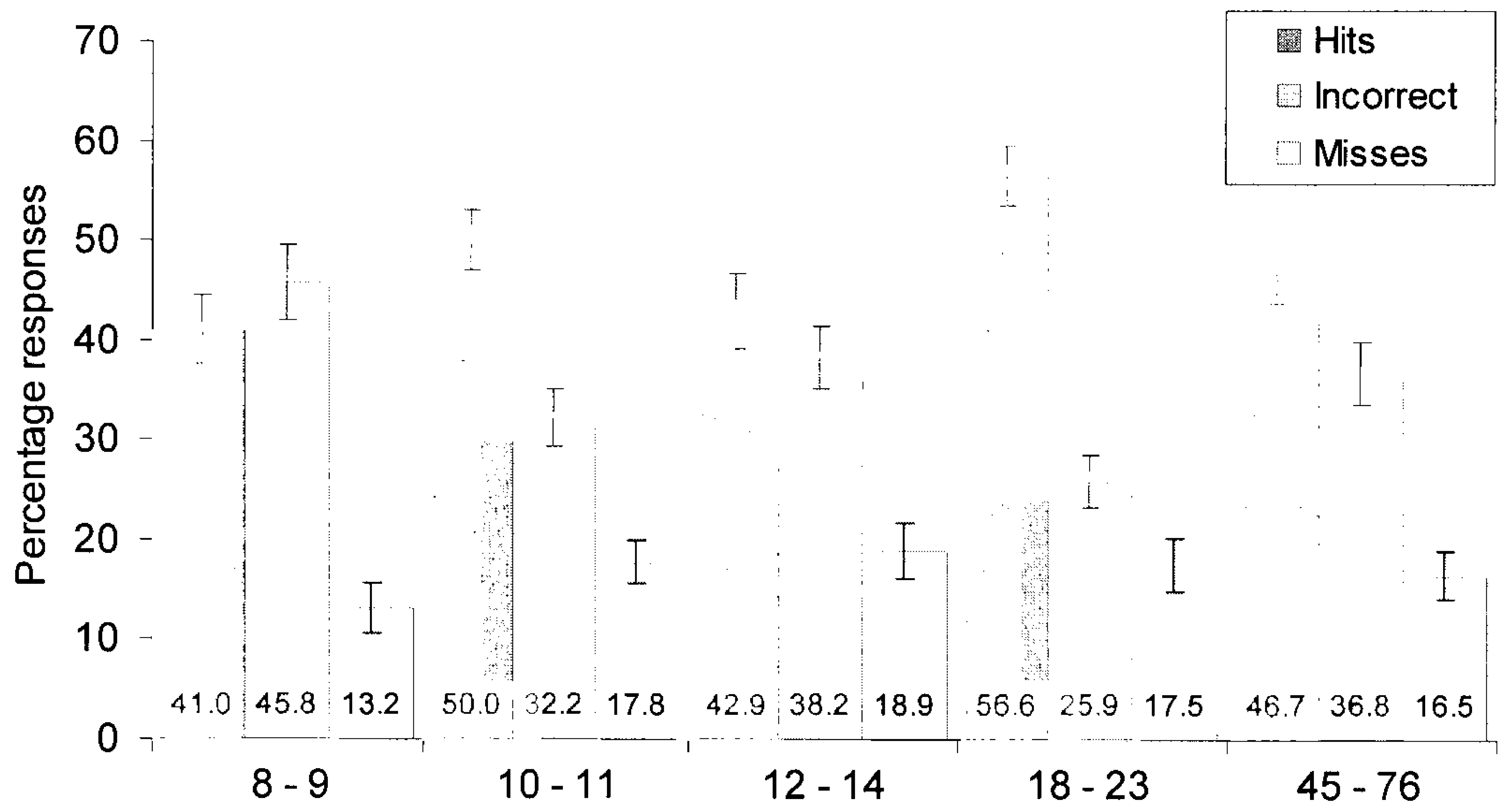



*Figure 4.2: Percentage of each type of target present response as a function of the five sub-divided age groups (error bars signify standard error of the mean)*

Three independent-measures one-way ANOVAs were conducted on the target present responses with the new age groups as the independent variable. There was a significant difference in the number of hits, $F\ (4, 270) = 3.58, p < .01$.

Games-Howell post-hoc tests found that two clusters emerged:

8 – 9    12 – 14    45 – 76    10 – 11    18 – 23

Both the 8 – 9 year-olds, and the 12 – 14 year-olds made significantly fewer hits than the young adults ($p < .05$).

The number of incorrect false positive selections also significantly differed, $F(4, 270) = 5.77$, $p < .001$. From this analysis, three post-hoc clusters emerged ($p < .05$).

```
8 – 9     12 – 14     45 – 76     10 – 11     18 – 23
-------------------------------
          -----------------------------------
                      -----------------------------------
```

The 8 – 9 year-olds made more incorrect responses than both the 10 – 11 year-olds ($p < .05$) and the young adults ($p < .001$). The 12 – 14 year-olds ($p < .05$), and marginally the older adults ($p = .056$) made more incorrect responses than the young adults. None of the other comparisons were significant ($p > .1$).

The final ANOVA found there were no differences in the number of incorrect misses by each age group ($p > .1$).

**4. 1. 2. 3. Error rates associated with specific targets[1]**

Examination of individual test items found a variation in the response accuracy. The proportion of hits in target present conditions ranged from 94.3% (Actor 17), to only 12.9% (Actor 02) correctly matching video actors when presented in arrays. Some targets were consistently incorrectly selected, both as false alarms when the target was not present, and as incorrect selections when the target was present. The most extreme examples of this were that 44.8% wrongly selected one actor's photograph (Actor 31) when a different actor was shown in the video (Actor 32).

A further distracter from one sub-group (Actor 13) was incorrectly selected more times than three of the targets when the videos depicting those targets were shown (Actor 08 on video, Actor 08 selected by 38% of participants; Distracter 13 selected by 54.2% --- Actor 09 on video, Actor 09: 20%; Distracter 13: 42.9% --- Actor 10 on video, Actor 10: 17.1%; Distracter 13: 25.7%). This demonstrates that error rates were not confined to particular actor pairs, but more than one actor could be mistaken for a specific distracter.

[1] Note: Photographs of all actors in each sub-group are displayed in Appendix A

There was also a variation in target absent conditions. Correct rejections of individual distracters ranged from 5.7% (Actor 23) to 77.1% (Actor 16). Indeed, 94.3% of participants incorrectly selected one distracter (Actor 25) when Actor 23 was shown on video.

### 4. 1. 3. Discussion

Experiment 4.1 confirmed the findings of previous published studies (e.g., Bruce et al., 1999; Henderson et al, 2001) and the specific predictions of this experiment in finding high error rates in a task in which participants were required to match video footage showing one actor against an array of high-quality photographs. Regardless of age group, only 4 out of the 420 participants correctly responded across all of the six trials. When the target was absent from the array, a distracter was selected in approximately 64% of trials. Furthermore, when targets were present in arrays, 50% of all selections were incorrect, in two-thirds of which participants again selected a distracter. The remaining one-third were false negative responses, reflecting incorrect beliefs that the target was absent from the array.

Experiment 4.1 also demonstrated that the matching of young adult faces at this task is inferior in children up to at least the age of 9, as the hit rate of 10 - 11 years-old was equivalent to that of adults. In addition, the performance of the 12 – 14 year-olds was closer to that of the 8 – 9 year-olds than to the 10 – 11 year-olds, representing a developmental dip in face matching ability. Similar reverses in early adolescence have been found in other face processing tasks (e.g., Carey et al., 1980; Diamond et al., 1983; Flin, 1980; 1985a; Soppe, 1986). In addition, marginal evidence was found for a decrement in matching performance in the older adult age group in comparison to the younger adults. Although these findings are consistent with those found in recognition tasks (e.g., O'Rourke et al., 1989; Searcy et al., 1999; Smith & Winograd, 1978) and age estimation (George & Hole, 1995), this appears to be the first involving adult face identification in the absence of any demands on memory.

It is probable that children are more exposed to and more interested in faces of their own age group, in comparison to adult faces. Indeed, expertise or knowledge-based accounts of the difference between children and adults at face processing tasks suggest that if a specific child is interested in particular topics or objects there recognition performance can be better than that of an adult (Davies, 1996b). Older and younger adults may also be respectively more experienced with faces of their own age. The more accurate performance by the young adults in Experiment 4.1 may therefore be due to the action of an own-age face processing advantage as a consequence of the requirements to match face of the same age, a finding demonstrated in previous published research (e.g., Chance et al., 1986; Fulton & Bartlett, 1991; George & Hole, 1995; Wright & Stroud, 2002). Indeed, research examining childhood face recognition tends to use images showing children of the same age as the participants, to avoid this bias (e.g., Brace, Hole, Kemp et al., 2001). However, the increased error rate found in the 12 – 14 year-olds is less easy to explain in terms of an age-related deficit as they were closer in age to the adult actors than to the younger children. Consequently, these results add support for a developmental dip and a general reduction in face processing skills at this age, not based on the age of the target faces.

It is not possible to distinguish between the pubertal and information processing theories for this developmental dip with the data produced from this experiment (e.g., Carey et al., 1980; Diamond et al., 1983; Soppe, 1986). Girls normally reach puberty earlier and therefore the dip would be expected to be found in the younger female children. However, there were no gender differences in performance, although this explanation cannot be discounted as pubertal status was not tested. It is also not possible to disregard the information processing model as no data was compiled as to the strategies used by children.

Previous face recognition and eyewitness research has found that children (Lindsay et al., 1997; J.F. Parker & Carranza, 1989; J.F. Parker & Ryan, 1993) and older adults (Smith & Winograd, 1978) are more likely to make more selections. This strategy generally leads to more false positive errors, especially if the target is absent. Experiment 4.1 found no age group differences in the selection of the 'not present' option suggesting that the greater accuracy by the young adults was not due to being more cautious in their responses, or to the younger children lacking the meta-cognitive skills to estimate the likelihood of a target to be present in an array or to understand the task demands.. Therefore, the increased error rate by the other age groups appears to represent an inferior ability to distinguish between the targets and distracters in arrays, as suggested by information processing accounts of face processing (Davies, 1996b). However, it is possible that the strategies used by children to select any item from an array may be more error-prone than those of adults. Therefore, a second, less demanding experiment was devised in which instead of an array of six photographs, a single target was presented requiring a 'same/different' response.

**Experiment: 4. 2**

In a forensic scenario and certainly in court, identification of an individual from CCTV would involve deciding whether a single defendant shown in a photograph was present in the video evidence rather than making decisions to arrays of faces. Henderson et al. (2001; Experiment 5), utilising a single-item identity-verification design demonstrated the difficulty in making this type of decision. Over one-quarter of participants (27.5%) incorrectly believed that high-quality images of two different actors were of the same person, with a further 45% believing that two images of the same individual were of different people. Although the image quality of the stills was much higher than those obtained from most CCTV systems, there was a time lapse between acquiring both images of the target actors and some of the errors may reflect appearance alterations. In addition, the results were based on the findings from only two target actors and two distracters. It may be that these faces are unrepresentative of the population, perhaps possessing physical peculiarities and therefore results may not generalise.

Experiment 4.2 was designed to utilise a similar methodology and design to that of Henderson et al. (2001). A subset of stimuli was selected from those used in Experiment 4.1 and the response scale allowed for an examination of decision confidence. Mixed results have been found when examining the accuracy-confidence relationship in matching studies. For instance, Henderson et al. (2001) found that although confidence was generally low, certainty in decisions was higher in target present conditions. However, confidence was not separately reported for correct and incorrect decisions and no statistical analyses were recorded. Bruce et al. (1999) also examined confidence in some of their series of face matching experiments, finding that correct decisions were associated with higher confidence than incorrect decisions, especially when the target was present.

Studies comparing adults and children's recognition confidence in eyewitness studies have also found conflicting results. Some have found no differences regardless of whether participants were accurate or not (e.g., J.F. Parker et al., 1986; Searcy et al., 1999). In contrast, J.F. Parker and Carranza (1989) found no differences in confidence between adults and 9 year-old children on a first lineup. However, the confidence of children was higher on a second lineup regardless of whether they were correct or not. Furthermore, J.F. Parker and Ryan (1993) found that the confidence of 9 year-old children was higher than adults, but only when making incorrect selections. There were no age group differences when selections were correct.

A novel group of volunteers was recruited for this experiment and hypotheses were specified based on the results of Experiment 4.1 and previous face matching and recognition studies. As such, it was therefore predicted that younger children; adolescents and older adults would perform less accurately at this task than younger adults. Furthermore, the confidence of children was expected to be higher than adults. However, no predictions were made as to confidence differences between adults and children of different ages.

### 4. 2. 1. Method

#### 4. 2. 1. 1. Participants

Four hundred and eighty-three (221 male, 262 female) participants were recruited for this experiment using the same procedures as in Experiment 4.1. Two hundred and twenty-eight were under the age of 18 ($M$ = 11.3, $SD$ = 2.5); the remainder were aged 18 and over ($M$ = 37.0, $SD$ = 11.9). None of the participants had taken part in the previous experiment or any of the pilot studies.

#### 4. 2. 1. 2. Materials

A sub-set of 12 of the photographs and videos of the actors in no disguise used in Experiment 4.1 were assembled for this study. Two video tapes were constructed showing footage of six different actors randomly ordered in each. The method of construction was such that if a particular actor was present in the first video, the actor matched highest for similarity of appearance by pilot participants would appear on the second. This created four *video presentation blocks* so that for instance, Actor A could be present in the video and photograph (target present – *block 1*); Actor A present in the video, with Actor B (the distracter rated highest in similarity) in the photograph (target absent – *block 2*); Actor A present in the photograph, with Actor B in the video (target absent – *block 3*); or Actor B could be present in both video and photograph (target present – *block 4*) (see Figure 3.2). These different experimental *video presentation blocks* were treated as a separate variable in all analyses to ensure this method did not bias the results. Responses were collected using the 8-point identity-confidence scale described in Section 3.9.

#### 4. 2. 1. 1. Design

This experiment employed a 2 (target presence) x 4 (age group) mixed design in which participants made single-item identity-verification decisions to a series of six different videos simultaneously presented with a single photograph. The repeated-measures factor was whether the photograph of the actor shown in the video was present or absent. Half of the six trials were target present, half were target absent, in which a photograph of a matched distracter was displayed alongside the video. The independent-measures factor was the age group of the participants. The dependent variables were unadjusted scores on the 8-point identity-decision scale as well as accuracy and confidence measures obtained from conversion of these scores.

#### 4. 2. 1. 4. Procedure

Conditions were similar to those in Experiment 4.1 in that participants viewed a series of six videos each showing a single actor. However, a single photograph approximately the height of the monitor screen was displayed instead of arrays. Participants used the identity-decision scale to record their responses.

### 4. 2. 2. Results

Participants were sub-divided into four age groups, one based on a median split of those aged 18 and above and the second based on a median split of those below the age of 18[2]. These divisions were slightly different than those in Experiment 4.1. A chi-squared test found that although numbers from each age group assigned to the four video presentation block conditions were not equal, this difference was not significant, *Pearson* $\chi^2(9, n = 483) = 8.13, p > .1$.

The video presentation blocks had been counterbalanced to ensure that all items appeared in the target present/absent combinations an equal number of times. However, it was apparent that some blocks were easier than others. Therefore, block, along with gender was initially entered into all the following analyses as further variables. However, the main effect of gender was not significant and nor were interactions between gender or block and age group ($p > .1$) and therefore data were collapsed across conditions. Identity-decision scale data were categorized for

[2] Note: All the analyses reported in Experiment 4.2 were replicated using a variety of different age group categories. For instance, in one analysis participants were sub-divided into eight age groups. However, the results of these further tests did not essentially differ from those included in this results section and are therefore not reported.

accuracy and Table 4.2 reports the mean unconverted scale scores, and the proportion of hits and false alarms for each age group.

#### 4. 2. 2. 1. Analysis of unconverted scale scores

A 2 (target presence) x 2 (age group) mixed design ANOVA was conducted on the unconverted scale scores. A high score on this scale (out of 8) indicates a strong belief that the video footage and the photograph depicted the same person. The main repeated measures effect of target presence was highly significant, $F(1, 479) = 595.09$, $p < .001$; as would be expected using this scale, mean target present scores were higher than target absent responses. However, the effect of age group, $F(3, 479) = 1.78$, $p > .05$; and the interaction, $F < 1$ were not significant.

*Table 4.2: Mean scale scores and proportion of hits and false alarms in target present and absent conditions for each age group*

| | | Target present | | | Target absent | | |
|---|---|---|---|---|---|---|---|
| Age group | n | Mean score | SD | Hits | Mean score | SD | False alarms |
| 8 - 11 | 122 | 6.13 | 1.54 | 0.76 | 3.39 | 1.58 | 0.33 |
| 12 – 17 | 106 | 5.75 | 1.48 | 0.72 | 3.40 | 1.48 | 0.32 |
| 18 – 38 | 129 | 5.66 | 1.36 | 0.70 | 3.35 | 1.25 | 0.30 |
| 39 + | 126 | 5.92 | 1.56 | 0.73 | 3.47 | 1.49 | 0.34 |
| Total | 483 | 5.87 | 1.49 | 0.73 | 3.40 | 1.45 | 0.32 |

#### 4. 2. 2. 2. Analysis of converted error rate data

Table 4.3 shows the frequency of errors made by participants over the six trials. The overall percentage error rate was 29.74% (1.78 errors out of 6). As can be seen, one participant made the maximum of six errors, with another 67 making no errors (13.9%).

A 2 (target presence) x 4 (age group) mixed design ANOVA was also conducted on the number of errors. This revealed a significant main effect of target presence, $F(1, 479) = 10.06$, $p < .005$, with participants making more target absent errors than target present errors. However, the main effect of age group, $F < 1$, and the interaction $F(3, 479) = 1.23, p > .1$ were not significant.

Furthermore, one-way ANOVAs conducted on the signal detection theory sensitivity, $A'$ and bias, $B''$ measures found no differences across age groups ($p > .1$).

*Table 4.3: Participant error frequency with percentages*

| | Frequency of errors made by individual participants | | | | | | |
|---|---|---|---|---|---|---|---|
| Errors | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| Frequency | 67 | 148 | 135 | 99 | 24 | 9 | 1 |
| Percentage | 13.9 | 30.6 | 28.0 | 20.5 | 5.0 | 1.9 | 0.2 |

### 4. 2. 2. 3. Confidence level data

Scale scores were converted for confidence level data as described in Section 3.9. Figure 4.2 illustrates the mean confidence levels across the age groups regardless of whether targets were present or absent.

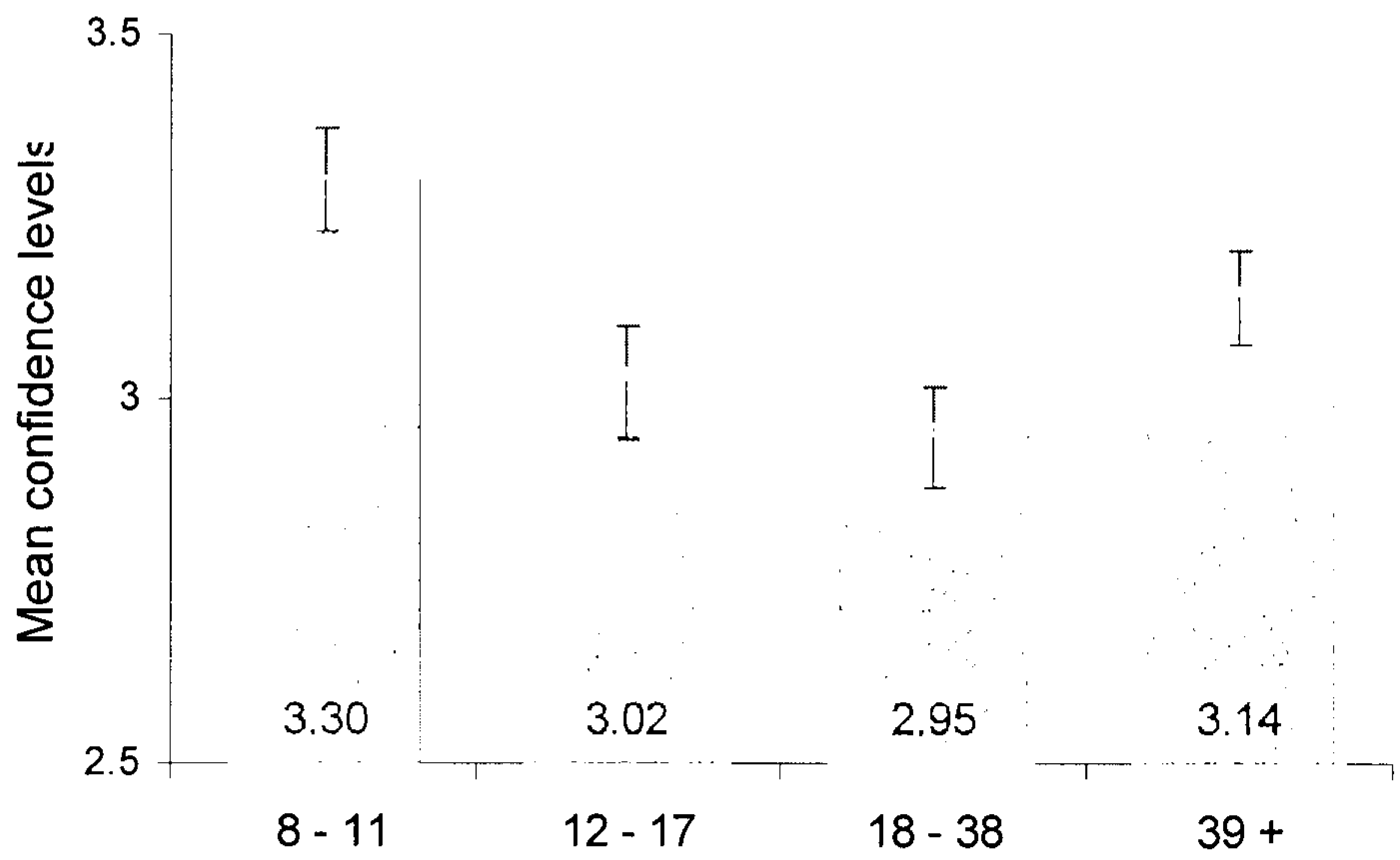



*Figure 4.3: Mean confidence levels (maximum = 4: minimum = 1) pooled across correct and incorrect responses as a function of age group (error bars denote standard error of the mean)*

To analyse confidence, two separate procedures were conducted as some participants made no errors in either target present ($n$ = 189), or in target absent conditions ($n$ = 141) and therefore did not contribute data to some cells. The first analysis, in which the target presence data was pooled, was designed to analyse the accuracy-confidence relationship across age groups. However, some participants ($n$ = 68) were still excluded as they were either 100% or 0% correct in all 6 trials. A 2 (accuracy) x 4 (age group) mixed design ANOVA conducted on these data revealed a significant main effect of accuracy, $F(1, 411) = 67.32$, $p < .005$; confidence in correct decisions was higher than when incorrect ($M = 3.23$, $SD = 0.65$ vs. $M = 2.84$, $SD = 0.95$). The main effect of age group was also significant, $F(3, 411) = 5.56$, $p < .01$. Games-Howell post-hoc tests found that this was due to higher confidence levels by 8 – 11 year-olds than both the 12 – 17 year-olds and the 18 – 38 year-olds whose confidence levels did not differ ($p < .05$).

The second procedure examined decision type ('same' or 'different') confidence levels as a function of age group. This allowed for a more powerful analysis as data from fewer participants ($n$ = 12) was excluded for failing to make at least one decision of both types. A 2 x 4 mixed design ANOVA conducted on these data found a non-significant repeated measures main effect of decision type, $F < 1$ and a non-significant interaction $F < 1$. However, the main effect of age group was significant, $F(3, 467) = 9.22$, $p < .001$. From this analysis, three Games Howell post-hoc clusters emerged.

```
8 – 11     39+      12 -17     18 - 38
--------------------
           --------------------
                    ---------------------
```

The age group differences were consistent with those reported above except that in this more statistically powerful analysis, the confidence levels of the older adults were additionally higher than that of the younger adults ($p < .05$).

### 4. 2. 2. 4. Error rates associated with specific targets[3]

As in Experiment 4.1, performance varied across both target present and target absent individual trials. In two of the target absent trials, almost two-thirds of participants wrongly believed that the video and the photograph showed the same person (Video: Actor 24, Photo: Actor 25 66% errors; Video: Actor 32, Photo: Actor 31 62% errors). Even in the most accurate target absent trial (Video: Actor 25, Photo: Actor 24), 8% of participants were incorrect. Furthermore, when Actor P was shown in both the photograph and video in target present trials, 65% of participants believed the two images were of different people. In the target present trial associated with the highest accuracy (Actor 32), 9% of participants were still in error.

### 4. 2. 3. Discussion

Experiment 4.2 again demonstrated that the identification of unfamiliar people, simultaneously shown in video and in a photograph is error prone. Across all age groups approximately 30% of decisions in the single-item identity-verification design of Experiment 4.2 were in error. Slightly more were made in target absent conditions (32%) than when targets were present (27%). Although no specific predictions had been made, there are two potential associated explanations for this effect. Firstly, the participants may have been using a liberal response criterion towards believing that the two images displayed the same person. This type of bias will tend to increase target absent errors, while reducing target present errors. Or, secondly, discrimination between the distracters and the target faces was problematical due to a strong resemblance in these images, effects possibly enhanced by the smaller quality and size of the video images. These alternative explanations are explored in later experiments within this thesis using the same images.

However, these video images were designed to meet Home Office specifications concerning the size of individuals shown in CCTV footage for accurate unfamiliar

---

[3] Note: Frontal photographs of all actors are displayed in Appendix E

identification purposes (Aldridge, 1989; 1994). Furthermore, they would almost certainly be deemed of 'sufficient' quality by a trial judge, in order that a jury could decide for themselves whether they believed the defendant was depicted in the absence of any other identification evidence. A replication of these effects by jurors being invited to make this type of decision when presented with a contemporary photograph of a defendant and CCTV footage of a culprit could obviously have serious consequences if images did not depict the same person.

However, in contrast to predictions these results were not consistent with previous published research that has found that younger children are inferior at this type of task (Bruce et al., 2000; Carey et al., 1980; Ellis, 1992). They were also not consistent with the results of Experiment 4.1 in which younger children, adolescents and older adults were worse than young adults in a matching task to arrays of photographs. Explanations for these differences are explored later.

Across all age groups, confidence levels were consistently higher when responses were correct than with incorrect responses, replicating the findings of previous matching studies (e.g., Bruce et al., 1999; Henderson et al., 2001). However, there were differences in confidence levels across the age groups. Younger children (8 – 11 year-olds) were more confident than the adolescent children (12 – 17 year-olds) and the younger adults, findings consistent with some published eyewitness research when memory is assessed (e.g., J.F. Parker & Carranza, 1989; J.F. Parker & Ryan, 1993). In addition, the older adults in this experiment were also more confident in their decisions than the younger adults.

One potential explanation for these differences may be as a result of the particular scale used. The overall proportion of responses at the extremes of the scale (1 & 8) by the younger children (65.2%) was considerably higher than by the other age groups (12 – 17: 44.5%; 18 – 37: 40.5%; 38+: 50.2%). This may indicate that the younger children did not understand the concept of using central marks on the scale, or don't have the meta-cognitive skills to judge the probability of an accurate response. Indeed, almost one-quarter (23.8%) of this group responded at the extremes of the scale across all six of their trials. In contrast, approximately one-in-ten participants in the other age groups were this restrictive in all their responses. Nevertheless, over three-quarters of the younger children did use a variety of response choices indicating that the concept of the scale was not too complex for the majority.

The findings of higher confidence levels in older adults are unlikely to be due a lack of understanding of the purpose of the scale. A similar proportion of this age group as the younger adults and older children responded at the scale extremes across all six trials, with over 90% using central choices on at least one of their trials. It would therefore appear that this difference was not the result of methodological factors but rather represents higher confidence in their ability to perform this task.

**4. 3. General discussion**

The primary aim of the two experiments reported in this chapter was to examine whether differences across age groups would be found in identity matching tasks. Experiment 4.1 demonstrated that young adults were consistently superior at this task when matching an actor of approximately the same age shown in video to their image within an array of six photographs. In contrast, the youngest children tested were more likely to select the wrong photograph even when the target was present. There was also evidence that older adults above the age of 45, and adolescent children between the ages of 12 and 14 were inferior at this task. These age group deficits are consistent with those found in young children and adolescents in previous published face matching and recognition studies (Bruce et al., 2000; Carey et al., 1980; Ellis, 1992). Similar differences have been found in some face recognition and eyewitness studies when comparing older and younger adults (e.g., O'Rourke et al., 1989; Searcy et al., 1999; Smith & Winograd, 1978). However, no previous published study appears to have found similar deficits in older adults in simultaneous face matching.

In contrast, no age group differences were found in the accuracy of responses in Experiment 4.2 in which matching decisions were made to a single photograph only. Potential explanations for the lack of effects are that the task was too simple or that there were too few trials and therefore no age group differences could be detected. However, in 8 out of the 24 combinations of paired stimuli (video/photograph) in this experiment, more than a third of the participants were in error (3 target present trials, 5 target absent), with at least 8% of participants in error in the most accurate single trial. This suggests there were no ceiling or floor effects. These results therefore suggest that children and adults are equally able to match the identity of faces when only two images are presented.

As such, the research reported here supports interrelated explanations for these effects. The first is that the own-age face processing advantage found in other research was operating, in the young adult participants when targets are presented within arrays of similarly-appearing distracters (e.g., Chance et al., 1986; George & Hole, 1995; Smith & Winograd, 1978; Wright & Stroud, 2002). All the target actors were approximately the same age as participants in this group and this would explain the inferior performance by the other age groups. In addition, the performance of the youngest children supports information processing accounts (e.g., Davies, 1996b), in that discrimination of highly similar faces is less effective leading to the identification of a distracter instead of the target. In contrast, in Experiment 4.2 when the task was to decide if a single actor shown in a photograph was depicted in the video, there was no requirement for participants to differentiate between different faces of a fairly similar appearance and therefore performance was constant across all age groups.

However, the performance of the 10 – 11 year olds was equivalent to the young adults and superior to both the 8 – 9 year-olds and 12 – 14 year-olds meaning that this cannot be the complete explanation. Indeed, this provides further evidence of a developmental dip in performance at the onset of adolescence. Furthermore, the increased confidence levels found in the older adults in Experiment 4.2 may represent an inappropriate belief in their ability to distinguish between faces. Taken with the results found in previous studies of a greater likelihood of making a selection (e.g., Smith & Winograd, 1978), even if the target is absent, this would suggest an inclination to a somewhat less cautious decision policy by older adults.

The secondary aim of the experiments reported in this chapter was to compare performance with previous studies examining face matching in adults. The error rates in Experiment 4.1 were comparable to those from previous studies in which participants were required to select a target shown in a close-up video still from an array of ten (Bruce et al., 1999) and eight (Henderson et al., 2001) facial photographs. Approximately 30% of responses were in error in the Bruce et al. study, with those by Henderson et al. considerably higher at 70% probably due to the lower quality of images. The array sizes in Experiment 4.1 were slightly smaller than in these two studies. However, approximately 53% of adult responses were incorrect. These errors were also probably higher due to using medium range videos, rather than the close-up images employed by Bruce et al. (1999).

The error rates using the single-item identity-verification design in Experiment 4.2 were comparable to those found by Henderson et al. (2001; Experiment 5). However, photographs of only two actors and two matched distracters were employed in that study. Therefore, the replicated findings in Experiment 4.2 using a much larger database of facial images add weight to the conclusions made by Henderson et al. as to the potential forensic implications of this type of task. Indeed, these findings suggest that care should be taken before presenting a jury with a photograph of a defendant and inviting them to compare the image with the culprit captured on CCTV footage. Nevertheless, the defendant would be present in person in court, potentially providing additional identification cues that might improve performance considerably. This issue is pursued further in Chapter 6.

The videos in the experiments reported in this chapter all depicted the actors wearing no disguise. Criminals often disguise themselves to avoid detection, with some stating that this would be a deliberate policy to avoid identification from CCTV (Loveday & Gill, 2003). Furthermore, video images from a crime scene may not always show the full face of the offender. The two experiments reported in Chapter 5 examine these factors, while retaining the design and the footage used in Experiment 4.2.

## Chapter 5: The simultaneous identity-matching of faces in disguise

### 5. 0. Introduction

A common *modus operandi* of criminals is the use of a disguise to obscure facial features and so reduce the likelihood of being identified. With increase in coverage and improvements in the quality of CCTV, convicted criminals in the UK have stated that they would be more likely to use a disguise, when carrying out their 'professional' activities (Loveday & Gill, 2003). They are probably correct to believe that they could evade prosecution in this manner. Empirical evidence has consistently demonstrated that transformations of facial stimuli impair recognition performance. This occurs with changes in lighting direction (Hill & Bruce, 1996), photographic negation (Kemp, Pike, White & Musselman, 1996), inversion (Valentine, 1988), variations in pose and expression (Bruce, 1982; Bruce et al., 1987), by the artificial masking of facial features in photographs (Young et al., 1985), and by the removal or addition of different disguises (e.g., wigs, hats, glasses: Diamond & Carey, 1977; hats: Flin, 1985a; dark glasses: Hockley et al., 1999; Metzger, 2001; glasses, wigs and beards: Patterson & Baddeley, 1977; clear glasses and beards: Terry, 1994). However, all of the above studies examined memory for faces. If CCTV images are obtained, it would be possible to directly match footage with photographs of the defendant. Therefore, the aim of the two experiments reported in this chapter was to examine performance in this context.

There is evidence that for the recognition of familiar people, the internal features of the face, such as the mouth, nose or eyes are more salient (Ellis et al., 1979; Young et al., 1985). However, a defendant would not be known to a jury or in most cases police officers; and external features such as hairstyle and face shape appear to be more important in the perception of unfamiliar faces (Bruce et al., 1999; Henderson et al., 2001; Young et al., 1985). For instance, Young et al. (1985; Experiment 1) instigated a simultaneous face matching study in which participants were required to decide if pairs of unfamiliar and familiar face images were of the same or of different people. Response times were faster when the external features of familiar faces were masked. The reverse was true for unfamiliar faces when internal features were masked. Further evidence for an unfamiliar face external feature advantage in a face matching design was found by Bruce et al. (1999; Experiment 4). Identification of a target from an array was more accurate if the internal features of the target were masked (73%) in comparison to the masking of external features (49%). Performance in the former condition was only marginally inferior to the matching of un-manipulated faces (84%).

O'Donnell and Bruce (2001) also demonstrated that hairstyle is particularly important in the processing of unfamiliar faces. The artificial manipulation of hairstyle, while being familiarised to photographed faces was noticed more often than changes to other facial features. Indeed, alterations to internal features such as the eyes were only discerned once participants had become more familiar with the faces. As it is extremely simple to manipulate or disguise hairstyle, the use of this as a primary identification cue is likely to induce errors in a forensic scenario.

To different extents, unless items such as motor cycle helmets or stocking masks are used, disguise will disrupt processing by *partly* obscuring internal or external features. However, only one study appears to have investigated the effects of disguise in a simultaneous matching task (Henderson et al., 2001; Experiment 3). Two 'culprit' actors were filmed twice performing a staged robbery, once wearing a hat and once in no disguise. Participants were presented with high-quality close-up CCTV stills and were required to select the same face from an array of eight photographs. Accuracy and confidence was higher when the target was undisguised. However, only two target actors were recruited in this study and it may be inadvisable to assume that findings can be generalised to a wider population. In addition, in court, a jury would be presented with images of a single defendant; not an array of faces as in this experiment.

**Experiment: 5. 1. Matching disguised faces on video with undisguised photographs**

Experiment 5.1 was therefore designed to examine whether the use of different disguises would impact on simultaneous matching ability from video footage to photographs. The single-item identity-verification design was retained from Experiment 4.2 and videos of the actors wearing a hat and a pair of dark glasses were utilised together with those depicting them in no disguise. These disguises were selected as being unobtrusive, dependent on the season of the year. All participants were unfamiliar with the actors and for target present conditions predictions were possible. The results of matching and recognition studies have consistently found that external features are more important in the perception of unfamiliar faces (e.g., Bruce et al., 1999; Henderson et al. 2001; Young et al., 1985), and therefore when the actor in the video was depicted wearing a hat, identification accuracy was expected to be impaired.

Predictions concerning the wearing of dark glasses were less clear as these partially obscure internal features only, which tend to be of lesser importance in unfamiliar face identification (e.g., Bruce et al., 1999). However, recognition studies have found that disguises primarily masking internal features such as clear or dark glasses also disrupt processing (e.g., Hockley et al., 1999; Metzger, 2001; Patterson & Baddeley, 1977; Terry, 1994) and the results of recognition and matching studies tend to display a similar pattern. Therefore, the wearing of glasses was expected to reduce performance but not to the same extent as a hat.

However, it was not possible to make firm predictions for target absent conditions. With unfamiliar faces, because of the importance of external features (e.g., Bruce et al., 1999; Young et al., 1985), especially hairstyle (O'Donnell & Bruce, 2001), misidentifications of two different faces are more probable if styling and length are similar. This might suggest that if one of the two faces in this experiment were depicted wearing a hat, performance might actually be improved as participants would be impelled to scrutinise the internal features of the actors making any differences more salient. In contrast, dark glasses would have no effect on matching using external features and performance would be predicted to be similar to that found when actors were in no disguise

It is also possible that when one face is shown wearing *any* disguise, participants might employ a more cautious criterion before inferring that identity is matched and therefore be biased to respond that the two images depict different people. This would result in improved accuracy in target absent conditions, while conversely reducing performance when targets were present. By analysing the signal detection measures derived from the accuracy data, as well as the identity-decision scale score data, it was therefore possible to detect any response bias of this type in this experiment.

### 5. 1. 1. Method

#### 5. 1. 1. 1. Design

This experiment utilised a 3 (disguise) x 2 (target presence) repeated measures single-item identity-verification design. The first factor was disguise, each participant viewed six videos, in which the actors were depicted equally often in no disguise, dark glasses or a hat. The second factor was target presence: half of the trials were target present; half were target absent. Therefore, each participant provided data for three target present trials. In each of these trials the video depicted the target actor in one of the three disguise conditions. Participants also provided data for three trials in which the target was absent, again there being one trial for each disguise condition. The dependent variables were unadjusted scores on the 8-point identity scale as well as accuracy and confidence measures obtained from conversion of the scores. As in Experiment 4.2, for presentation purposes, the videoed actors were sub-divided into four different *video presentation block* conditions. All possible combinations of these conditions were presented to an equal number of participants across the blocks, and were fully counterbalanced and randomised.

### 5. 1. 1. 2. Participants

Six hundred (279 male, 321 female) adult participants took part in this study (Mean age = 33.3, *SD* = 12.2). All were visitors to the Science Museum, London and were recruited using the same method as described in Experiment 4.1. None had taken part in any previous experiment or any of the pilot studies.

### 5. 1. 1. 3. Materials and procedure

The photographic stimuli for this study were the 24 high-quality three-quarter view black-and-white facial photographs of the actors whose videos had been utilised in Experiment 4.1. For target present conditions, the actor shown in the photograph was also depicted in the simultaneously presented video footage. For target absent conditions a matched distracter was presented. Videos of the actors taken at the same time in either no disguise, a hat or in dark glasses (Figure 5.1) were utilised for the different disguise conditions (See: Section 3.3.2). The procedure was the same as in Experiment 4.2.

*Figure 5.1: Disguise conditions in Experiment 5.1*

### 5. 1. 2. Results

The results were analysed in five different ways, unadjusted scale analyses, as a function of errors using discrete data from the scale; signal detection theory measures of sensitivity and bias and confidence analysis (Section 3.10). All post-hoc tests were conducted using the Bonferonni correction.

### 5. 1. 2. 1. Unadjusted scale score analysis

The overall means of the scale scores are presented in Figure 5.2 plotting the effects of disguise on responses in target present and absent trials. Of most interest in this analysis was the main effect of disguise as it provides an indication of whether disguise biased responses and is a more sensitive measure of bias than $B^{//}$ in this context. However, both are reported in this chapter. Mean scale scores that are higher than 4.5 indicate a bias towards responding that the two images are of the same person. The opposite is true for scores below 4.5.

A 2 (target presence) x 3 (disguise) repeated-measures ANOVA conducted on the unadjusted scale scores revealed a significant main effect of target presence, $F(1, 599) = 731.86$, $p < .001$, mean target present scores were as expected higher than target absent scores ($M = 5.62$, $SD = 1.34$ vs. 3.37, $SD = 1.31$). The main effect of disguise was also significant, $F(2, 1198) = 20.33$, $p < .001$. Post hoc tests found that scores in the no disguise condition ($M = 4.81$, $SD = 1.55$) were significantly higher than in the glasses ($M = 4.44$, $SD = 1.60$, $p < .001$) and hat conditions ($M = 4.23$, $SD = 1.49$, $p < .001$). There was also a marginally significant trend towards scores being higher in the glasses than in the hat condition ($p = .072$).

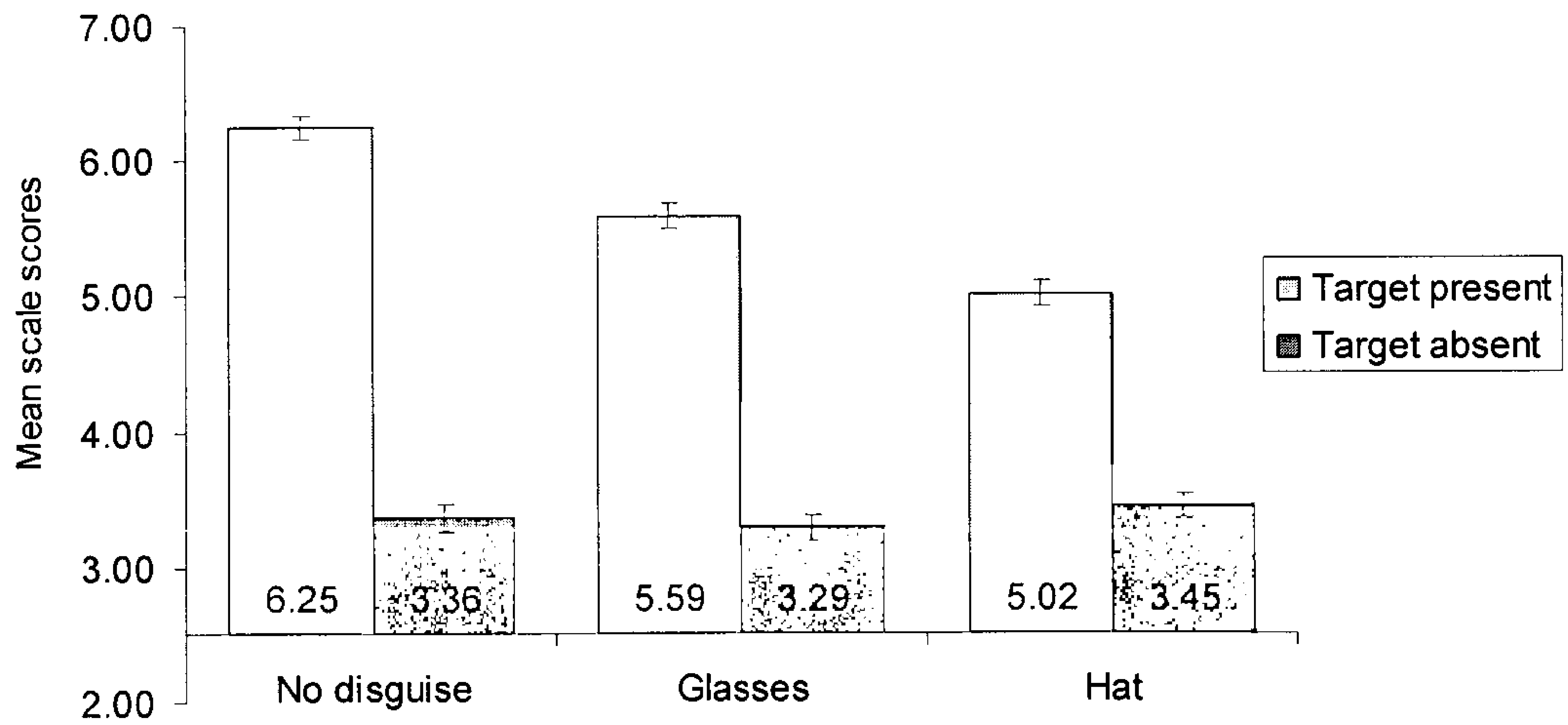



*Figure 5.2: Mean scale scores as a function of target presence and disguise (error bars denote standard error of the mean)*

The interaction was also significant, $F(2, 1198) = 25.57$, $p < .001$. Simple effects analyses found that in target present conditions scores in each of the disguise conditions were significantly different from one another, $F(2, 1198) = 46.56$, $p < .001$. Unplanned comparisons revealed that scores were highest when the actors were in no disguise and lowest when they were shown wearing a hat ($p < .001$). There were no differences across disguises when targets were absent, $F < 1$.

**5. 1. 2. 2. Accuracy analyses**

Identity decision scale data were also categorised for accuracy. Eighty-nine (14.8%) participants made correct decisions across all six trials, whereas only one participant made six incorrect responses. Figure 5.3 shows the percentage of errors as a function of target presence and disguise.

A 2 (target presence) x 3 (disguise) repeated-measures ANOVA conducted on this data found that the main effect of target presence was not significant, $F < 1$. However, the main effect of disguise was significant, $F(2, 1198) = 11.88$, $p < .001$; post-hoc tests found that more errors were made in the hat condition (34.7%, $SD = 36.3$) than in the no disguise (25.6%, $SD = 31.5$; $p < .001$) and glasses conditions (29.1%, $SD = 32.8$, $p < .05$). The no disguise and glasses conditions did not differ ($p > .1$).

This interaction was also significant, $F(2, 1198) = 14.19$, $p < .001$. Simple effects analyses found that in target present conditions the error rates in each of the disguise conditions were significantly different from one another, $F(2, 1198) = 25.99$, $p < .001$, with unplanned comparisons revealing that the fewest errors were made in the no disguise condition, followed by the glasses condition and the most were made in the hat condition ($p < .005$). In contrast, there was no difference when targets were absent, $F < 1$.

Accuracy rates for each disguise condition were combined to calculate the signal detection theory non-parametric sensitivity ($A'$) and bias ($B''$) statistics. A one-way repeated measures ANOVA on the sensitivity data across disguises was significant, $F(2, 1198) = 11.88, p < .001$. Post hoc tests found that sensitivity was significantly higher in the no disguise condition than the glasses condition and the hat condition ($A^{/} = .74$; $SD = .31$ vs. $.71$; $SD = .33$ & $.65$; $SD = .36$ respectively, $p < .05$). However, $A^{/}$ did not differ between the glasses and no disguise conditions ($p > .1$).

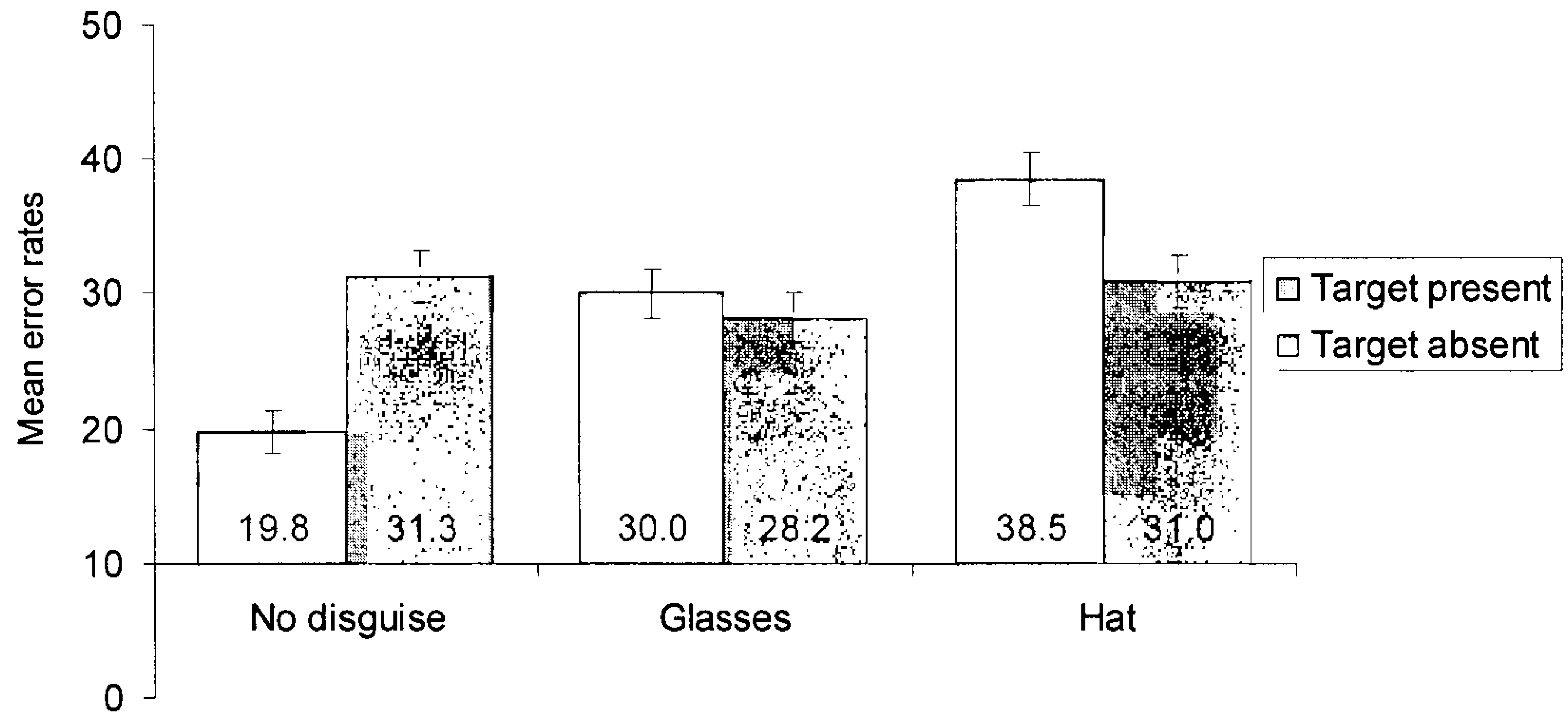



*Figure 5.3: Percentage error rate as a function of target presence and disguise (Error bars denote standard error of the mean)*

A similar ANOVA conducted on the bias $B^{//}$ showed a significant effect of disguise, $F(2, 1198) = 6.30, p < .005$. Post hoc tests found that $B^{//}$ was significantly higher in the no disguise condition than the hat condition ($B^{//} = .28$; $SD = .25$ vs. $.24$; $SD = .25$; $p < .05$), showing that responses were less conservative when a hat was worn than in the no disguise condition. The glasses condition ($B^{//} = .26$; $SD = .25$) did not differ from the other disguise conditions ($p > .1$).

### 5. 1. 2. 3. Confidence analysis

Three separate analyses were conducted on converted confidence level data. These could not be combined into a single analysis as numbers would have been low in some cells. The first analysis examined the relationship between accuracy and confidence. The mean confidence level for correct responses was 2.96 ($SD = 0.65$), for incorrect responses this was 2.51 ($SD$ = 0.87). A repeated-measures t-test conducted on this data found that this difference was significant, $t(509) = 11.23, p < .001$[1].

A similar test examining confidence in response type found that 'different' decisions ($M$ = 2.87, $SD$ = 0.75) were associated with higher confidence than 'same' decisions ($M$ = 2.81, $SD$ = 0.70), $t(593) = 2.04, p < .05$[2].

A final 3 (disguise) x 2 (target presence) repeated-measures ANOVA conducted on the confidence level data revealed a significant main effect of disguise, $F(2, 1198) = 40.96, p < .001$; post hoc tests found that confidence was significantly higher when actors were in no disguise ($M$ = 2.99, $SD$ = 0.77), than in glasses ($M$ = 2.84, $SD$ = 0.79) which was also significantly higher than when they were wearing a hat ($M$ = 2.65, $SD$ = 0.82, $p < .001$ all comparisons). However, neither the main effect of target presence, $F < 1$, nor the interaction were significant, $F(2, 1198) = 2.79, p > .05$.

**5. 1. 2. 4. Calibration between confidence and accuracy**

From the converted confidence data it was possible to measure the calibration between confidence and accuracy across the three disguise conditions. This involved plotting the *subjective* probability of a correct response, or the confidence of participants in decisions against the *objective* probability of being correct as measured by the accuracy data. A perfect calibration between these measures would occur if identification performance was 100% when rated confidence was 4, and at chance (50%) when rated confidence was 1. A large deviation from these values would indicate a weak calibration between accuracy and confidence (Brewer, Keast & Rishworth, 2002).

---

[1] Note: Some data was missing from this analysis as a number of participants made either no errors ($n$ = 89), or were 100% incorrect ($n$ = 1).

[2] Note: Data was missing as some participants responded with one type of decision only ($n$ = 6).

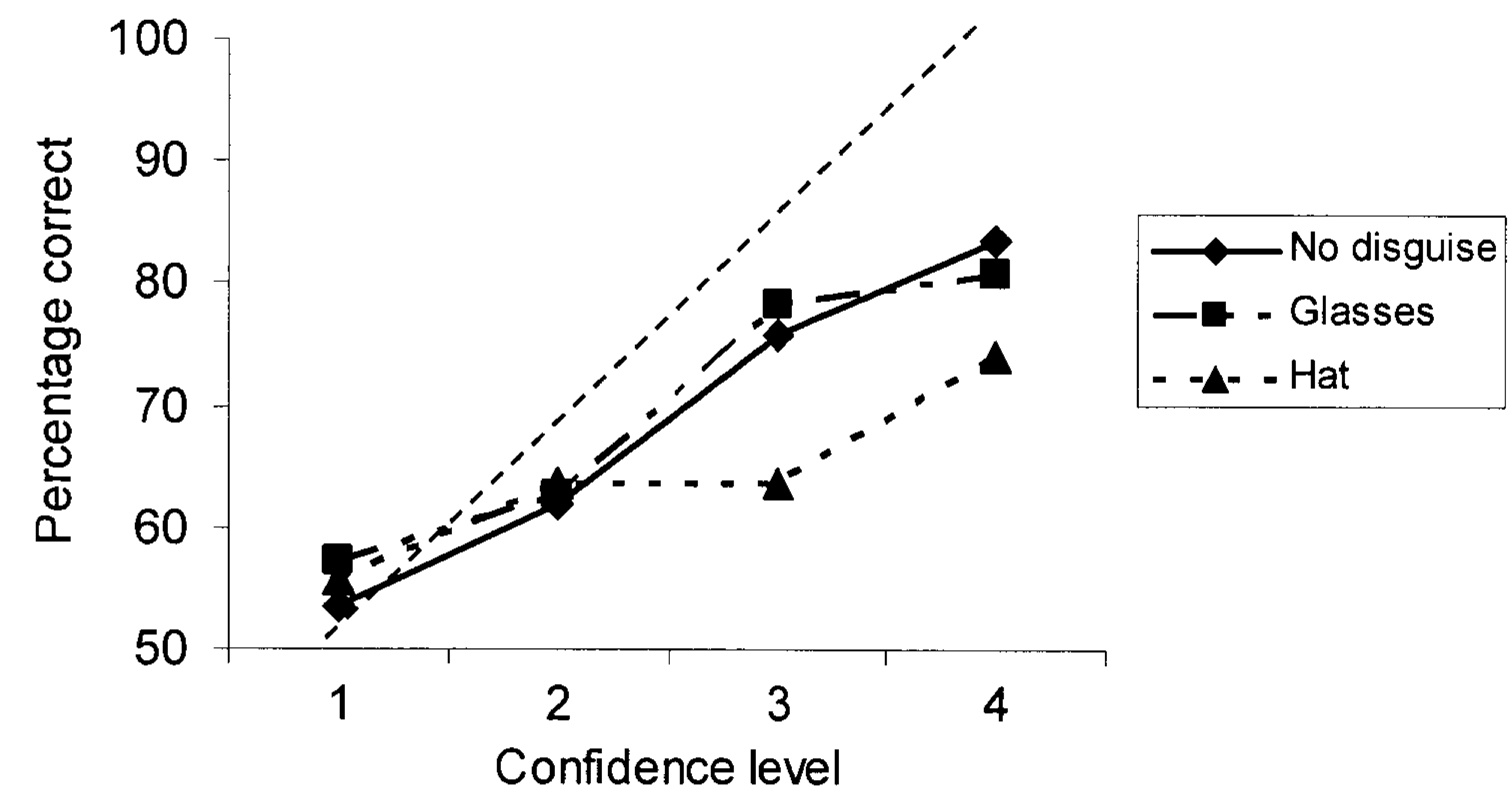



*Figure 5.4: Calibration between accuracy and confidence. The mean percentage of correct responses at each of the four confidence scale intervals for the three disguises (the diagonal line indicates perfect calibration)*

Figure 5.4 depicts the percentage accuracy of responses at the four levels of converted confidence for each disguise condition. The dashed diagonal line indicates the line of perfect linear calibration function for these data. When participants were less confident in their responses (intervals of 1, 2 or 3), the data closely matched this function when the actor was depicted in either no disguise or in dark glasses. However, participants tended to be over-confident with higher response ratings, particularly when the actor was wearing a hat as the calibration function can be seen to deviate from the diagonal line. Indeed, only 74.1% of participants were accurate in the hat condition when responding with the highest confidence levels.

### 5. 1. 2. 5. Individual item error rates

There was again a wide range in accuracy to individual items. In target present conditions the maximum number of errors associated with actors in no disguise was 52% (Actor 11), glasses 68% (Actor 24) and hat 64% (Actor 11). In contrast, no errors were made in target present conditions involving some of the other actors. In target absent conditions the highest number of errors in the no disguise conditions were associated with Actor 37 (mistaken for Actor 36 in video: 72%), glasses Actor 37 (36 in video: 84%) and hat, equally Actors 37 and 23 (Actors 36 and 22 in video respectively: 60%). No target absent pairing was associated with 100% correct responses.

### 5. 1. 3. Discussion

The results of Experiment 5.1 were consistent with previous published research (e.g., Bruce et al., 1999; Henderson et al., 2001), as well as the findings of the experiments reported in Chapter 4 in that the simultaneous matching of facial images is error prone. Overall error rates were approximately 30%. Performance was also adversely affected by the use of disguise, in particular the hat. However, disguise specifically disrupted the ability to identify that two images of the *same* person actually portrayed that same person. Error rates were highest in target present conditions when the actors in video were shown wearing a hat, followed by glasses and lowest when they were shown in no disguise. The results from this matching study therefore correspond with those examining the effect of disguise on facial recognition (Diamond & Carey, 1977; Flin, 1985a; Hockley et al., 1999; Metzger, 2001; Patterson & Baddeley, 1977; Terry, 1994) and on simultaneous matching performance (Henderson et al., 2001).

In contrast, disguise was found to have no additional impact on the accuracy of decisions when the two images depicted two different people in target absent conditions. Approximately 30% of all responses were incorrect, indicating a wrongful belief that the same person was shown in both images. A perfunctory analysis of the results would therefore suggest that disguise only affects target present decisions. However, the results from the unadjusted scale score main effect of disguise analysis (and to a lesser extent the non-parametric bias $B^{//}$ statistic) indicate that regardless of target presence there was a bias towards responding that the target was absent when the videoed actor was wearing a disguise. Furthermore, there was marginal evidence that wearing a hat increased this bias in comparison to wearing glasses. The operation of this conservative bias with disguised faces had the effect of reducing the accuracy of target present responses while also reducing the number of target absent errors. As such, if the conservative criterion adopted by participants when confronted by images showing the actors in a disguise was also found in a forensic scenario, there would be a lower likelihood of an innocent suspect being wrongly identified as the offender in video.

Similar to the results of Experiment 4.2, correct decisions were associated with higher levels of confidence than incorrect decisions. However, whereas no differences in confidence were found in Experiment 4.2 between 'same' and 'different' decisions, in Experiment 5.1, confidence in 'different' decisions was higher than in 'same' decisions. This may be due to the actors being disguised in videos in this experiment. Indeed, confidence was highest in the no disguise condition and lowest in the hat condition regardless of whether the same actor was present or not. Nonetheless, from the calibration data, it can be seen that many participants responded with inappropriate high levels of confidence, particularly in the hat condition. Those responding with lower levels of confidence had a more realistic appraisal of their ability to perform this task.

In Experiment 5.1 the photographs were three-quarter close-up images, taking up almost the entirety of the monitor screens, providing detailed high-spatial-frequency facial information. In contrast, the relatively low-spatial-frequency video footage showed the actors full bodies with no facial close-ups. The pattern of results found in Experiment 5.1 may therefore to some extent be due to the discrepancy in spatial frequencies. However, reducing the resolution of the photographs to match the video footage in Experiment 5.1 would have been unlikely to have impacted on performance. Liu et al. (2003a) found that if the spatial resolution of test photographs was higher than a video image, matching and recognition of unfamiliar faces was more accurate than if the photograph resolution was consistent with the low-quality video. However, the effects found in Experiment 5.1 may also have been the result of a discrepancy in image size. Therefore, Experiment 5.2 was instigated to examine whether consistent effects would be found if the high resolution photographic images depicted the actors in disguise instead of the videos.

### Experiment: 5. 2. Matching disguised faces in photographs with undisguised videos

This experiment was again designed to examine the impact of disguise on matching performance. However, instead of the video footage showing the actors in the three disguise conditions; disguise was manipulated in the photographs only. In contrast to Experiment 5.1 all videos depicted the actors in no disguise.

### 5. 2. 1. Method

#### 5. 2. 1. 1. Participants

Six hundred (296 male, 304 female) adult visitors to the Science Museum, London were recruited for this study (Mean age = 33.9, *SD* = 12.0) using the same method as described in Experiment 4.1. None had taken part in any previous experiment or pilot study.

#### 5. 2. 1. 2. Design, materials and procedure

The design and procedure replicated Experiment 5.1 except the photographic stimuli for this study were the 72 high-quality three-quarters view black-and-white facial photographs. These were of the same 24 actors taken in the each of the three disguise conditions described in Section 3.4. The video images showed the same actors in no disguise only.

### 5. 2. 2. Results

#### 5. 2. 2. 1. Unadjusted scale scores

Figure 5.5 shows the mean unadjusted scale scores as a function of target presence and disguise. A 2 x 3 repeated measures ANOVA conducted on these data found a significant main effect of target presence, $F(1, 599) = 838.84$, $p < .001$; with scores higher in target present than target absent conditions ($M = 5.78$, $SD = 1.31$ vs. $M = 3.38$, $SD = 1.40$). The effect of disguise was also significant, $F(2, 1198) = 14.11$, $p < .001$; paired comparisons found that mean scores were significantly lower in the hat condition ($M = 4.31$, $SD = 1.64$; $p < .01$) than in the no disguise ($M = 4.81$, $SD = 1.54$) and glasses conditions ($M = 4.61$, $SD = 1.58$; $p < .01$). The no disguise and glasses conditions did not differ ($p > .1$).

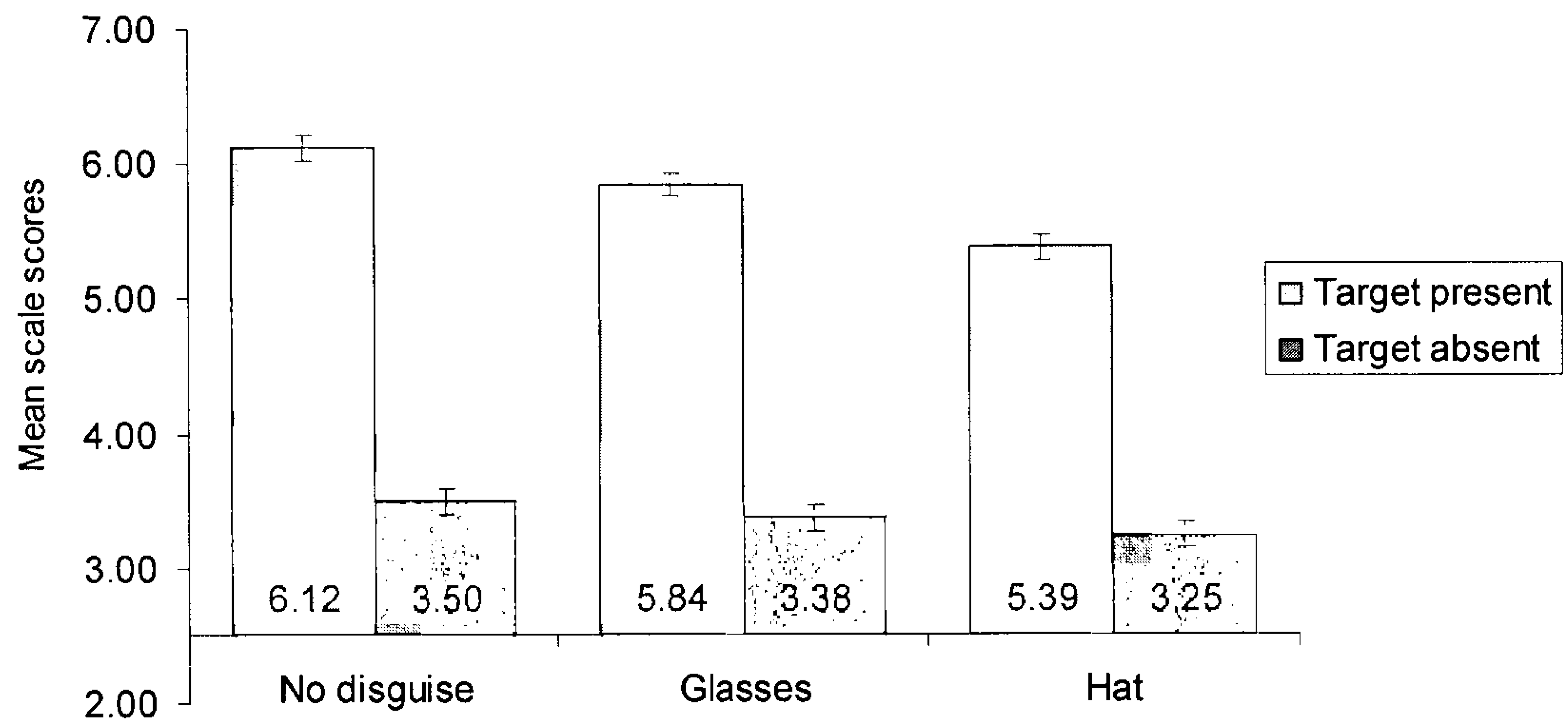



*Figure 5.5: Mean scale scores as a function of target presence and disguise (Error bars denote standard error of the mean)*

The interaction was also significant, $F(2, 1198) = 3.62$, $p < .05$. Simple effect analyses revealed a significant disguise effect in target present conditions, $F(2, 1198) = 16.27$, $p < .001$; scores were higher when the actors were depicted in no disguise and glasses than when wearing a hat ($p < .005$). There were no differences between the no disguise and glasses conditions when the targets were present ($p > .1$) and the simple effects of disguise were not significant when targets were absent, $F(2, 1198) = 1.78$, $p > .1$.

#### 5. 2. 2. 2. Error rate analysis

Scale scores were converted for analyses of accuracy and are presented in Figure 5.6 as a function of target presence and disguise. None of the 600 participants made 6 errors on all of their 6 trials. However, 102 (17.0%) made no errors at this task. A

2 x 3 repeated measures ANOVA conducted on these data found a significant main effect of target presence, $F$ (1, 599) = 5.89, $p$ < .05; with more errors in target absent ($M$ = 30.4%, $SD$ = 27.1) than in target present conditions ($M$ = 27.1%, $SD$ = 25.4). The main effect of disguise was not significant, $F$(2, 1198) = 2.41, $p$ > .05.

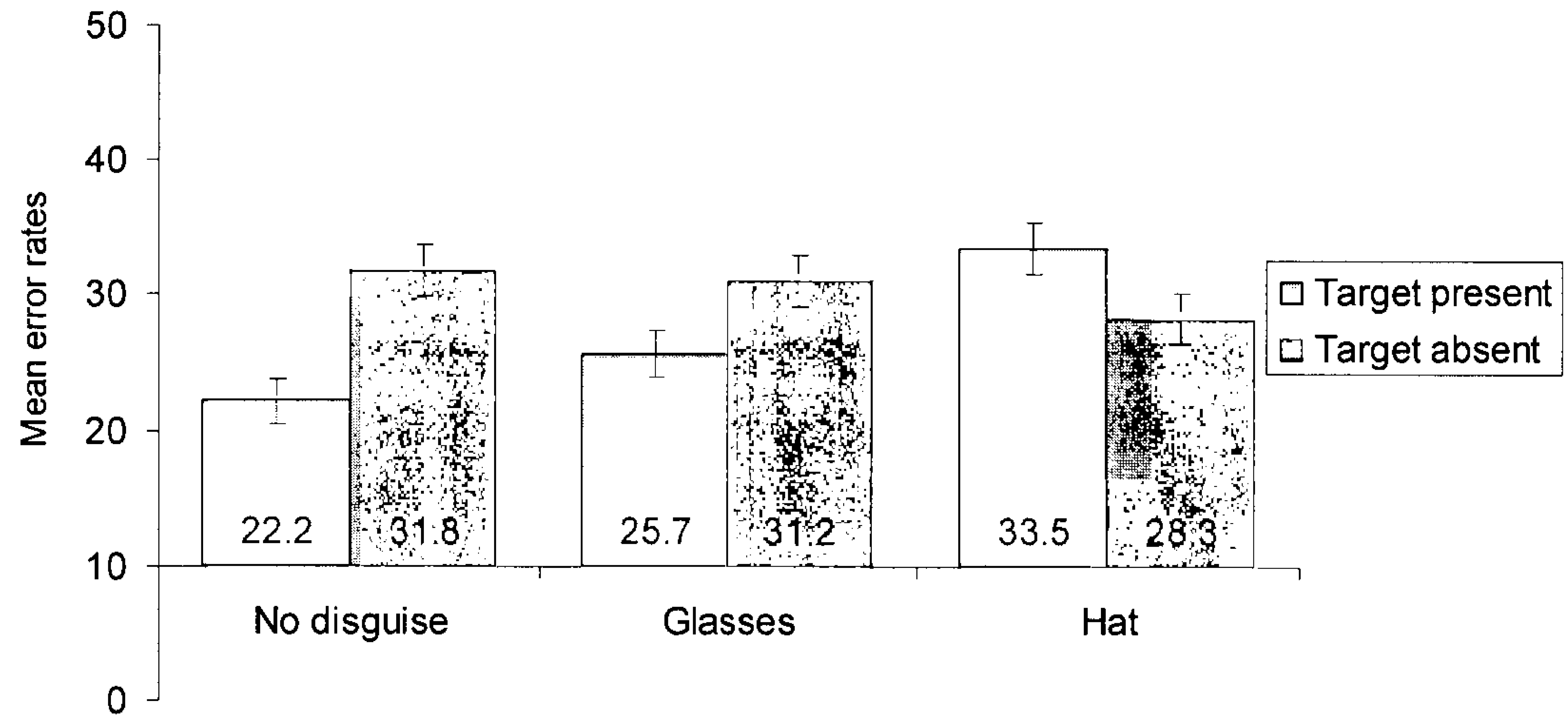



*Figure 5.6: Percentage error rate as a function of target presence and disguise (error bars denote standard error of the mean)*

However, the interaction was significant, $F$(2, 1198) = 8.42, $p$ < .001. Simple effect analyses revealed a significant effect of disguise in target present conditions, $F$(2, 1198) = 10.29, $p$ < .001. Paired comparisons revealed that significantly more errors were made in the hat condition than in the no disguise ($p$ < .001) and glasses conditions ($p$ < .05) The no disguise and glasses conditions did not differ ($p$ > .1). There were also no significant differences in the number of errors made in each disguise condition when targets were absent ($F$ > 1).

Accuracy rates for each disguise condition were combined to calculate the sensitivity ($A'$) and bias ($B''$) statistics. Sensitivity was highest in the no disguise condition ($A'$ = .73; $SD$ = .33), next highest in the glasses condition ($A'$ = .72; $SD$ = .33) and lowest in the hat condition ($A'$ = .69; $SD$ = .32). However, a one-way repeated measures ANOVA on these data was not significant, $F$(2, 1198) = 2.41, $p$ > .05.

In contrast, a similar ANOVA conducted on the bias $B''$ data was significant, $F(2, 1198) = 4.68, p < .01$; post hoc tests found that a significantly more liberal criterion was used in the no disguise condition than the hat condition ($B'' = .28$; $SD = .25$ vs. $.23$; $SD = .25$; $p < .05$). The mean value for glasses ($B'' = .26$; $SD = .25$) did not significantly differ from the other conditions ($p > .1$).

### 5. 2. 2. 3. Confidence analysis

Scale scores were converted in order to conduct three separate analyses of confidence. The first examined confidence as a function of accuracy. A repeated-measures t-test found that correct responses were associated with higher confidence (M = 3.02, $SD = 0.65$) than incorrect responses ($M = 2.51$, $SD = 0.93$); $t(497) = 12.68, p < .001$[3].

A similar test was conducted to compare confidence in decision type. This revealed that confidence levels in 'same' ($M = 2.88$, $SD = 0.71$) and 'different' ($M = 2.85$, $SD = 0.79$) decisions did not significantly differ, $t(593) = 0.84, p > .1$[4].

A 2 (target presence) x 3 (disguise) repeated measures ANOVA was also conducted on the converted confidence level data, revealing that the main effect of target presence was not significant ($p > .1$). However, the main effect of disguise was significant, $F(2, 1198) = 16.10, p < .001$. Post hoc tests found that scores in each of the disguise conditions were significantly different from each other ($p < .05$). Confidence was highest in the no disguise condition ($M = 2.98$; $SD = 0.81$), followed by the glasses condition ($M = 2.86$; $SD = 0.79$); and was lowest in the hat condition ($M = 2.76$; $SD = 0.85$).

The interaction was also significant, $F(2, 1198) = 8.47, p < .001$; simple effects analyses revealed that confidence scores showed the same pattern reported above for target present conditions, $F(2, 1198) = 24.08, p < .001$, with confidence highest

[3] Note: Data was missing from this analysis as a number of participants made no errors ($n = 102$).
[4] Note: Data was missing as some participants responded with one type of decision only ($n = 6$).

in the no disguise and lowest in the hat conditions. However, there was no difference in confidence between disguise conditions when targets were absent, $F(2, 1198) = 1.01, p > .1$.

### 5. 2. 2. 4. Calibration between confidence and accuracy

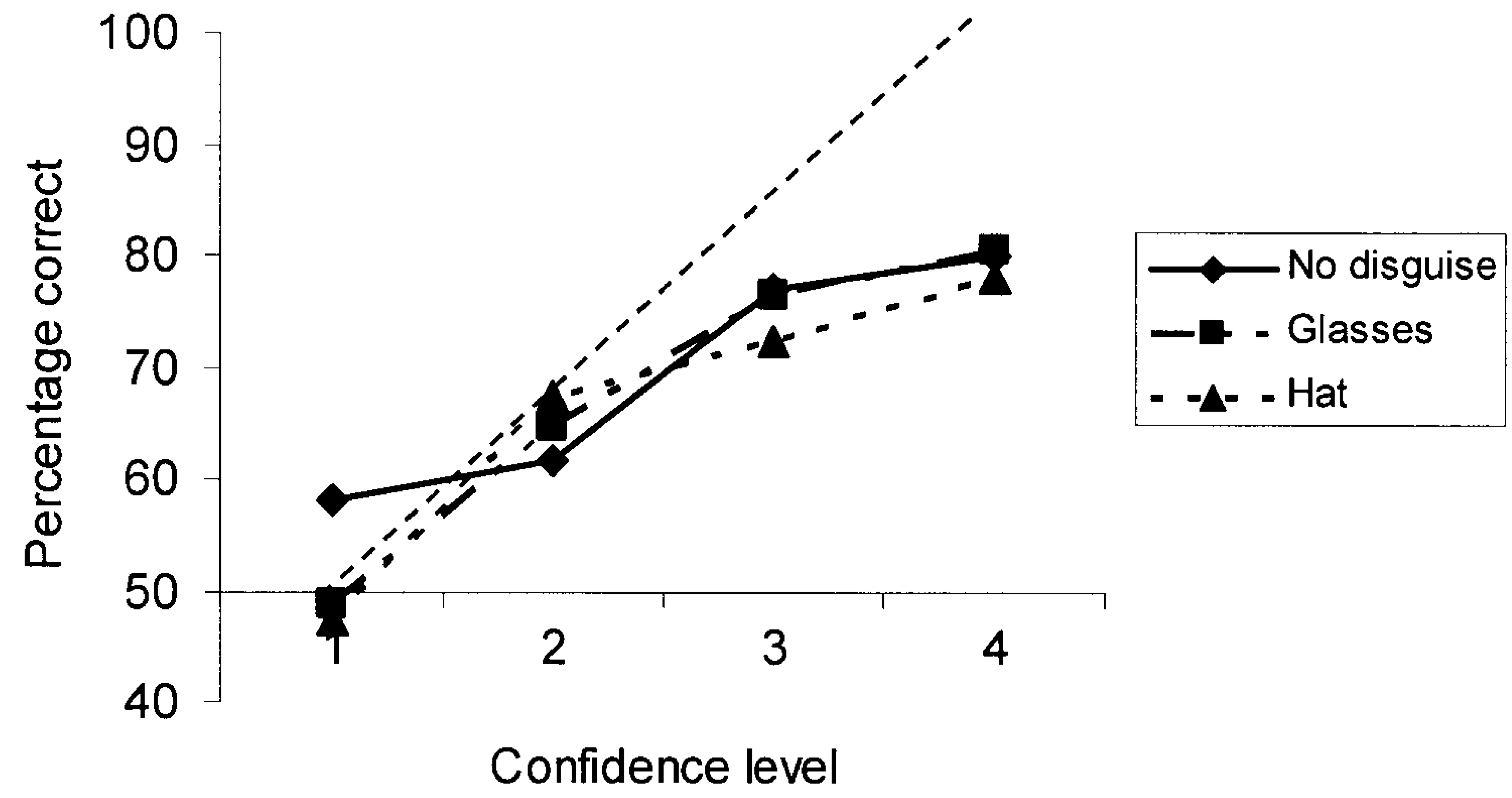



*Figure 5.7: Calibration between accuracy and confidence. The mean percentage of correct responses at each of the four confidence scale intervals for the three disguises (the diagonal line indicates perfect calibration)*

The calibration curves for each disguise condition are presented in Figure 5.7. For both conditions in which actors wore a disguise at confidence level = 1 (hat = 47.3%, glasses = 48.7%) accuracy is close to chance levels of 50%. However, accuracy is approximately 80% at confidence = 4 in all conditions, far lower than the expected 100%, indicating that participants were over-confident in decisions when responding with the highest levels of confidence. Although there is a diagonal line indicating a relationship between confidence and accuracy in each condition, these values illustrate that the association is less than perfect in this experiment.

### 5. 2. 2. 3. Individual item error rates

Examination of the number of errors associated with different stimuli again found a range of scores across items. In target present conditions the maximum number of errors associated with actors in no disguise was 68% (Actor 18), glasses 72% (Actor 11) and hat 72% (Actor 11). In contrast, there were no errors for some of the other actors. In target absent conditions the highest number of errors in the no disguise conditions were associated with Actor 36 (Actor 37 in video: 72%), glasses Actor 01 (Actor 02 in video: 84%) and hat, Actor 32 (Actor 31 in video 68%).

### 5. 2. 3. Discussion

In Experiment 5.2, approximately 29% of participants made errors in deciding whether two different images depicted the same person, or whether they depicted two different people. However, these effects were mediated by the use of a disguise and were to some extent similar to those of Experiment 5.1. When actors were depicted wearing a hat, a conservative response bias had the effect of increasing the number of errors in target present conditions only. There were also slightly more errors when actors were depicted wearing glasses, but these did not significantly differ from when they were shown in no disguise. However, consistent with Experiment 5.1 there were no differences in error rates across disguise conditions when images of two different people were displayed in target absent trials.

## 5. 3. General discussion

In court, juries may be presented with a contemporary photograph of a defendant, taken at approximately the same time as CCTV images of an incident, in order that they may be invited to conclude that both images are of the same person (Attorney General's Reference, No. 2 of 2002, 2003). In the two experiments reported in Chapter 5, photographs and video footage of the same actors were obtained within 24 hours. Across both experiments when the actors were shown in no disguise, approximately 20% of participants made incorrect identification decisions in target present conditions and a further 30% made inaccurate responses in target absent trials. However, error rates were higher when the same actor was depicted in one of the two images wearing a disguise, especially if it was a hat. These findings are consistent with previous studies and again supports previous research by demonstrating that decisions of this type are unreliable if faces are unfamiliar (e.g., Bruce et al., 1999; Bruce et al., 2001; Burton et al., 1999; Henderson et al., 2001). Moreover, in both experiments these figures represent average values; error rates were considerably higher in some individual trials.

Thus it appears that the wearing of a hat reduces accuracy in target present conditions, mainly by inducing a more cautious conservative response. These results are consistent with studies finding a disadvantage in matching and recognition when targets are depicted in a hat (e.g., Diamond & Carey, 1977; Flin, 1985a; Henderson et al., 2001).

When the image of a person shown in dark glasses is less distinct, as in the medium range images shown in Experiment 5.1 a similar bias appeared to be operating, in that participants were again more cautious in their responses. However, this bias did not occur with the high-spatial-resolution close-up images showing the same faces in dark glasses in Experiment 5.2. Indeed, the pattern of results in Experiment 5.2 in the dark glasses condition was more similar to when both images were shown in no disguise, suggesting that participants disregarded the loss of internal facial detail in the close-up views.

In both experiments, accuracy was associated with higher confidence in decisions, similar to that found in Experiment 4.2 and in previous matching studies (e.g., Bruce et al., 1999). As such, confidence was higher when decisions were correct than when they were incorrect. In the experiments reported in Chapter 5, confidence in decisions was also associated with the disguise worn by the targets. However, these results were again not consistent. In Experiment 5.1, confidence was highest when actors were in no disguise and lowest when they were wearing a hat regardless of whether the same actor was depicted or not. However, in Experiment 5.2, the same effect was only found when targets were present. There were no differences in confidence when targets were absent.

The experiments reported in this chapter have implications in terms of the use of CCTV evidence presented in court, particularly if a jury is invited to compare the images with a contemporary undisputed photograph of the defendant. If the defendant is the person in footage and is depicted wearing a disguise, it is probable that individual jurors may be more cautious in inferring identity and subsequently could be more inclined to deliver a not guilty verdict. This increased caution was evident in these experiments when the actor was shown in a hat regardless of the spatial resolution of the images. Furthermore, in the hat condition, decision-confidence was also lower, and if replicated by jurors would probably again result in a not guilty verdict, if the standard of proof, 'beyond reasonable doubt' was applied. The findings from Experiment 5.1 suggest that a similar result would occur if the same defendant was shown wearing dark glasses in low-spatial-resolution images. However, if depicted in dark glasses in high-spatial-resolution images as in Experiment 5.2, the likelihood of a guilty verdict appears to be the same as if no disguise had been worn.

In contrast, there were no differences in accuracy rates across disguises in target absent conditions. This suggests that if a defendant was wrongly accused of a crime and CCTV evidence was presented in order to induce jurors to make a positive identification, the likelihood of a guilty verdict would not depend on whether the real offender was depicted in a disguise or not.

However, regardless of the ability of participants at this task when the matching stimuli were facial photographs, a defendant would be present at their own trial. Furthermore, the videos and photographs of the same person used in Chapter 5 were acquired within 24 hours of each other. Unless arrested at the scene of the crime, a surveillance image and a comparison photograph of a defendant would rarely be captured on the same day.

Therefore, the experiments reported in the Chapter 6 again employed the single-item identity-verification design. However, actors were live in person and identification sessions took place some time after the simulated video surveillance images were acquired.

# Chapter 6: Live actors and simultaneous identity-matching to video

## 6. 0. Introduction

The results of the experiments reported in Chapters 4 and 5, as well as previous published research (e.g., Bruce et al., 1999; Bruce et al., 2001; Davies & Thasen, 2000; Henderson et al., 2001), have demonstrated that matching faces shown in video with facial photographs is error prone. And yet, CCTV evidence is sometimes presented in a courtroom, with juries invited to conclude that the defendant is shown in the footage (e.g., *R v Dodson and Williams*, 1984). In these cases, it is possible that if presented with balanced prosecution and defence evidence, an individual juror may be inclined to base their verdict solely on whether they perceive there is a likeness or not to the culprit shown on video. However, in court the defendant will be present in person; and the studies above did not examine matching to live actors, only to photographs. A still facial photograph cannot provide information about a variety of potential identifying features such as height, weight or gait that may be of use when the target is physically present. Individuals may also have idiosyncratic posture and characteristic expressions, not apparent in posed static images.

In highly degraded images, movement has been found to aid familiar face recognition (e.g., Knight & Johnston, 1997; Lander et al., 1999; Lander & Chuang, 2005). With exceptions (e.g., Christie & Bruce, 1998; Liu et al., 2003a), movement has also been found to facilitate unfamiliar facial identification (e.g., Pike et al, 1997; Schiff et al., 1986). Gait recognition from extremely impoverished images such as dynamic point-light displays of highly-familiar people is also possible (Cutting & Kozlowski, 1977; Loula et al., 2005). Indeed, Stevenage, Nixon and Vince (1999) demonstrated that it is possible to learn and later identify the idiosyncratic stride patterns of previously-unfamiliar people from this type of display. Furthermore, human gait can be individuated by automatic recognition algorithms (e.g., Nixon & Carter, 2004). Due to these factors, Egan et al (1977) argue that identifications with live targets will always be more reliable than those using photographs. However, mixed results have been obtained from recognition studies, with some finding an advantage with live actors (e.g., Egan et al., 1977), some no differences (Shepherd et al., 1982) and others finding a photograph advantage, particularly with children (Dent, 1977; Dent & Gray, 1975 cited in Dent, 1977; Dent & Stephenson, 1979; Peters, 1991).

All of these studies measured recognition memory. Only one published study appears to have examined simultaneous face matching using live actors. Kemp et al. (1997) found that experienced supermarket cashiers were unable to detect the majority (64%) of 'fraudsters' presenting photograph credit cards of another person matched for appearance. However, the photographs were 2-$cm^2$, possibly too small to distinguish detailed facial features.

The three experiments reported in this chapter were therefore designed to examine identity matching using a forensically-relevant scenario, during which participants had to decide whether an actor present in person was depicted in video footage. In Experiments 6.1 and 6.2, the performance of jury-age adults was examined. In Experiment 6.3 the performance of children was compared to that of adults.

Bruce et al. (1999) reported wide variations in identification performance to individual stimuli in their photograph to video matching studies. Similar variations in performance across different faces were also found in the experiments in Chapters 4 and 5. In target present conditions, high error rates to some faces may have been due to specific photographs not providing a good resemblance to the same target in the video. In target absent conditions, the variation in performance was possibly due to different levels of facial similarity across actors affecting discrimination. To ensure that these effects were not an artifact of using photographs, analyses compared identification performance to the different live actors in all the following experiments.

**Experiment: 6. 1. Matching live actors with video images**

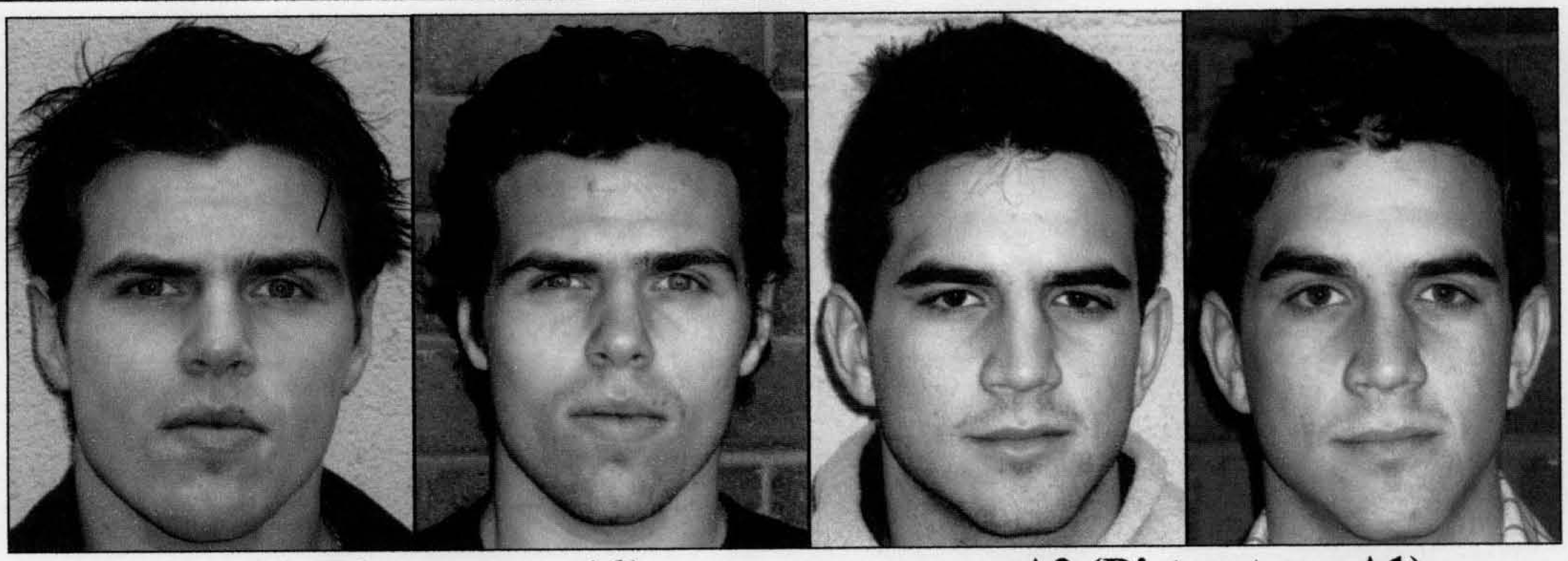
Experiment 6.1 was specifically designed to simulate the decision faced by a jury when the identity of somebody caught on CCTV is in dispute. Namely, is the defendant the person in the video? Participants viewed the footage and were asked to decide if the live actor walking into the room and standing by a monitor screen was shown in the video. This procedure was specifically designed to encourage participants to utilise gait cues. In cases that involve more than one defendant, jurors may be asked to make judgements concerning identification of two or more people, Therefore, participants took part in either one or two separate sessions, to test whether performance in a second judgement, concerning a different 'defendant' was influenced by their first. All sessions took place approximately three weeks after the footage was taken.

It has been suggested that corporeal identifications will always be more accurate than those made to photographs, due to additional cue availability (e.g., Egan et al., 1977). This would suggest that performance in this experiment would be better than to those previously reported with photographs. However, the photographs used in Experiments 4.1 to 5.2 were captured within 24 hours of the video footage, and any advantage from identifications being live, could be offset by appearance changes because of the increased time interval. Therefore, apart from expectations that a proportion of participants would again make identification errors, no specific predictions were made concerning actual performance for this experiment.

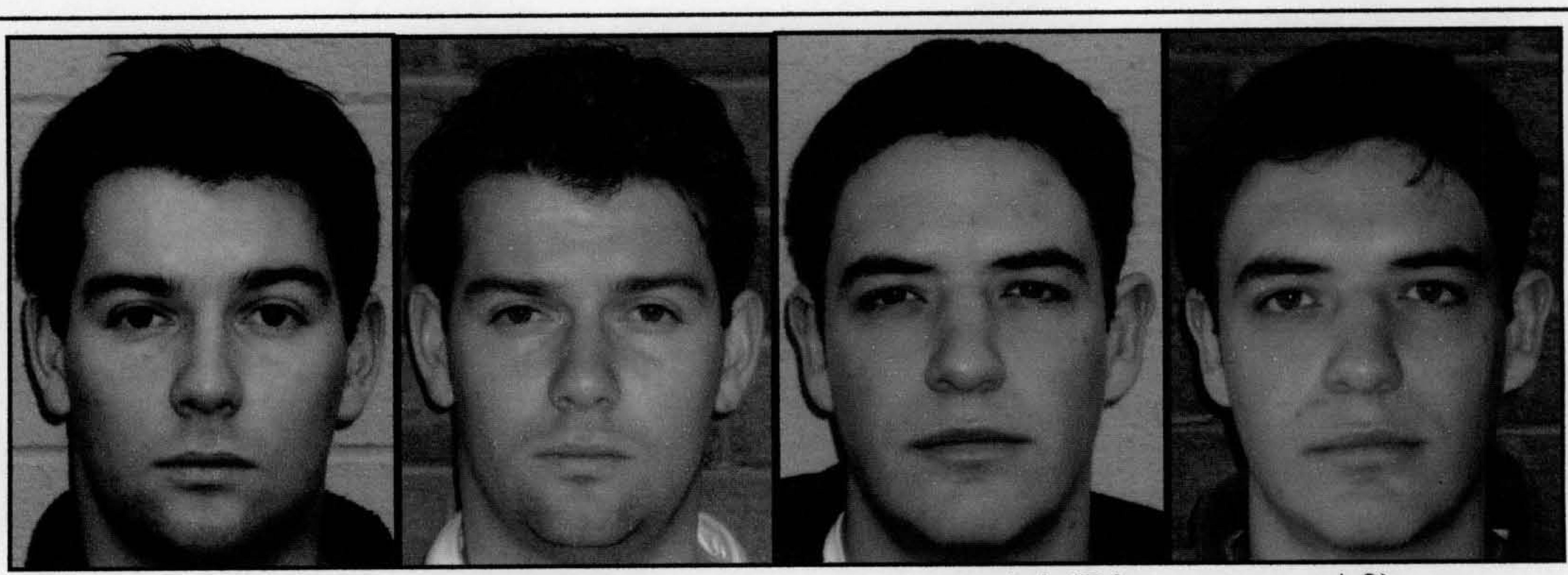
**6. 1. 1. Method**

**6. 1. 1. 1. Participants**

One hundred and ninety-eight (44 male, 154 female) adult visitors, staff and student volunteers (Mean age = 25.3, *SD* = 7.5) at Goldsmiths College, University of London contributed data to this study. None had taken part in any of the previous experiments or pilot studies.



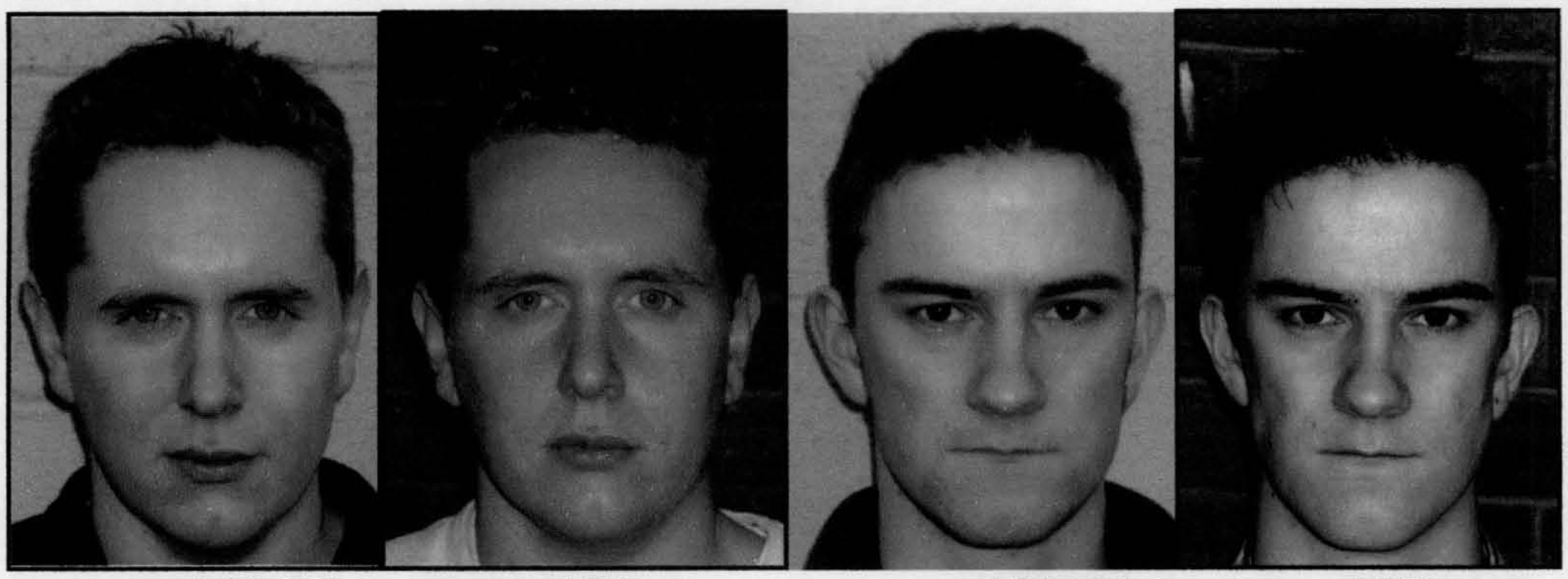
**A1 (Distracter = A2)** **A2 (Distracter = A1)**



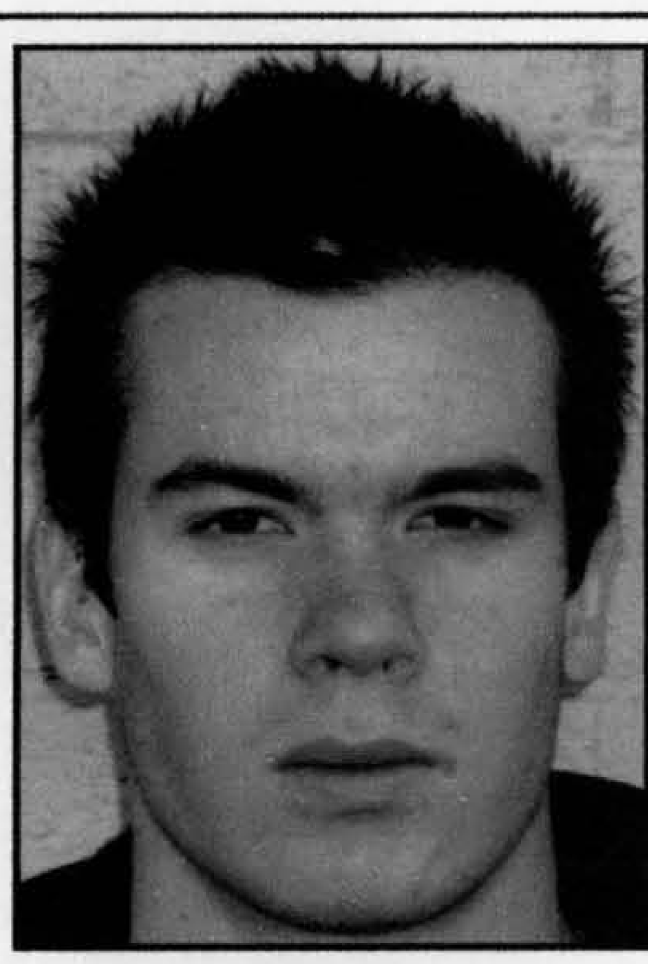
**A3 (Distracter = A4)** **A4 (Distracter = A3)**

**A9 (Distracter = A10)** **A10 (Distracter = A9)**

*(See page 143 for explanation of this figure)*



**A08 (Distracter = A09)** **A33 (Distracter = A17)**



**A17 (No distracter)**

*Figure 6.1: Two photographs of each actor recruited for Experiment 6.1. The first photograph was taken at the time of filming (left); the second was taken at the time of identification session (right). The letter associated with the distracters depicted in target absent conditions is also listed. Note: Actor 17 was unable to attend an identification session and therefore no distracter is indicated. However, the video taken of this actor was used in the target absent trial for Actor 33.*

#### 6. 1. 1. 2. Actors

The nine volunteer actors described in Section 3.1 were recruited for this experiment. Full-face photographs of each actor taken both at the time of the video session and later at the identification session are shown in Figure 6.1. The distracter for each live actor in target absent conditions was the video of the actor who was selectively paired as bearing the closest resemblance by the most pilot participants using full-length images taken at the same time as the videos.

#### 6. 1. 1. 3. Design

This experiment used a single-item identity-verification design. Participants attended one, or two identification sessions and in each, viewed a video clip and were required to decide using the identity-decision scale whether a physically present actor was depicted in the footage. Eight different actors took part in identification sessions, each attending one target present and one target absent session. For target present conditions a video of the actor taken approximately three weeks previously was shown. For target absent trials the video was of a matched distracter. The dependent variables were unadjusted scores on the eight-point identity-decision scale as well as converted accuracy and confidence measures.

**6. 1. 1. 4. Procedure**

Participants attended one, or two ($n$ = 136) video identification sessions. However, videos and actors in a second session were always different to that of the first. Trials were conducted either in a lecture theatre (see details below for the single trial conducted under these conditions), computer laboratory lecture groups, or in a series of single participant private viewings, although no specific data was collected to in order to differentiate between responses made under these two different conditions. Videos, replayed at least three times, were presented either on a large display screen (note: this occurred in one trial only, see below, $n$ = 51) or on individual monitors with viewers seated approximately 1m from the screen. In all trials the actors walked into the room and stood still with arms folded, keeping a neutral expression while the video was playing. The maximum distance from the actor to any participant in any trial was approximately 6 metres.

In the single target absent trial using the large display screen, the footage depicted Actor 17 while Actor 33 was present in the room. The size of the videoed actor on the screen in this trial would obviously appear larger to observers than when appearing on the computer monitors in all other sessions. Although evidence from face recognition and matching experiments involving perspective transformations suggest that this would have little effect on identification performance (Liu & Chaudhuri, 2003), the results from this specific trial are discussed separately later.

The actors wore different clothing from the videos and were blind to which footage was being shown. Participants viewed the video sequence and responded using the identity-decision scale, their belief as to the identity of the actor. The experimenter asked the participants to keep quiet throughout the sessions and observed them closely to ensure that none were collaborating in any manner. Full performance feedback was provided at the end of each session.

### 6. 1. 2. Results

#### 6. 1. 2. 1. Attendance at one or at two live identification sessions

Two analyses were conducted to examine whether there was a difference in responses between first and second sessions attended. A 2 (target presence) x 2 (session attended) ANOVA conducted on the scale data found a highly significant main effect of target presence, $F(1, 330) = 269.77$, $p < .001$; with target present scores higher than target absent scores ($M = 5.87$; $SD = 1.91$ vs. 2.59; $SD = 1.74$). However, both the main effect involving the number of sessions factor, $F(1, 330) = 3.25$, $p > .05$, and the interaction, $F(1, 330) = 1.91$, $p > .05$ were not significant.

A two-way chi-squared analysis was also conducted using the converted accuracy data. The two factors were session attended with two levels (first/only session or second session) and accuracy (correct or incorrect). This also revealed no significant difference in the accuracy of first and second sessions (20.7% vs. 17.6% respectively), *Pearson* $\chi^2(1, n = 334) = 0.482$. Therefore, in all subsequent analyses, the data from the two sessions were pooled.

#### 6. 1. 2. 2. Unconverted scale data

Unlike all other experiments reported in this thesis, the actors were not precisely paired so that if for instance Actor X acted as a distracter for Actor Y, Actor Y would not necessarily be the distracter for Actor X, meaning it was not possible to combine target present and target absent data into a single analysis. Therefore two separate one-way ANOVAs were conducted on the unconverted scale data. The first revealed no difference in target present scale scores ($M = 5.87$, $SD = 1.91$) across the eight actors, $F(7, 151) = 1.38$, $p > .1$. However, a second ANOVA conducted on the target absent scale scores across live actors was significant, $F(7, 167) = 12.96$, $p < .001$. Games-Howell post-hoc tests revealed four clusters ($p <$

.05), with low scores indicating a confident correct belief that the actor physically present was not depicted in the video footage:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Live actor: | A33 | A02 | A10 | A03 | A08 | A04 | A09 | A01 |
| Distracter: | A17 | A01 | A09 | A04 | A09 | A03 | A10 | A02 |
| M = | 1.37 | 1.86 | 2.33 | 3.00 | 3.10 | 3.40 | 3.90 | 4.38 |
| *SD* = | *(0.60)* | *(1.39)* | *(1.20)* | *(1.90)* | *(2.00)* | *(1.78)* | *(1.71)* | *(1.54)* |

### 6. 1. 2. 3. Accuracy data

Scale scores were converted for accuracy. A series of Binomial tests were conducted to examine whether performance differed from chance in each of the individual actor identification sessions and the results of these, together with the percentage error rate to each live actor in target present and in target absent conditions, are shown in Figure 6.2.

The proportion of errors in target present conditions varied dependant on each actor, ranging from 100% accurate performance in sessions involving Actor 01, whereas when three other actors were present in video, performance did not significantly differ from chance (Actors 02; 04 & 10). Indeed, 9 out of 24 (37.5%) participants incorrectly believed that Actor 02 was not shown in video footage. A similar variation was found across target absent trials. No errors were made when Actor 33 was present in person and Actor 17 was shown in video[1]. However, performance was at chance in four trials, with 7 out of 16 participants (43.8%) incorrectly responding that Actor 01 was shown in footage, when the video depicting Actor 02 was playing. Overall, there were more errors in target present conditions (22.0%) than in target absent conditions (17.1%), although a chi-squared analysis found that this was not significant, *Pearson* $\chi^2(1, n = 334) = 1.26, p > .1$.

[1] Note: this identification session was the only trial conducted using videos projected onto a large screen. However, it is unlikely that these particular results were due to these different conditions as the two actors involved had been paired the least the number of times by pilot participants, suggesting as can be seen in Figure 6.1.1 that their dissimilar appearance was unlikely to result in identification errors.

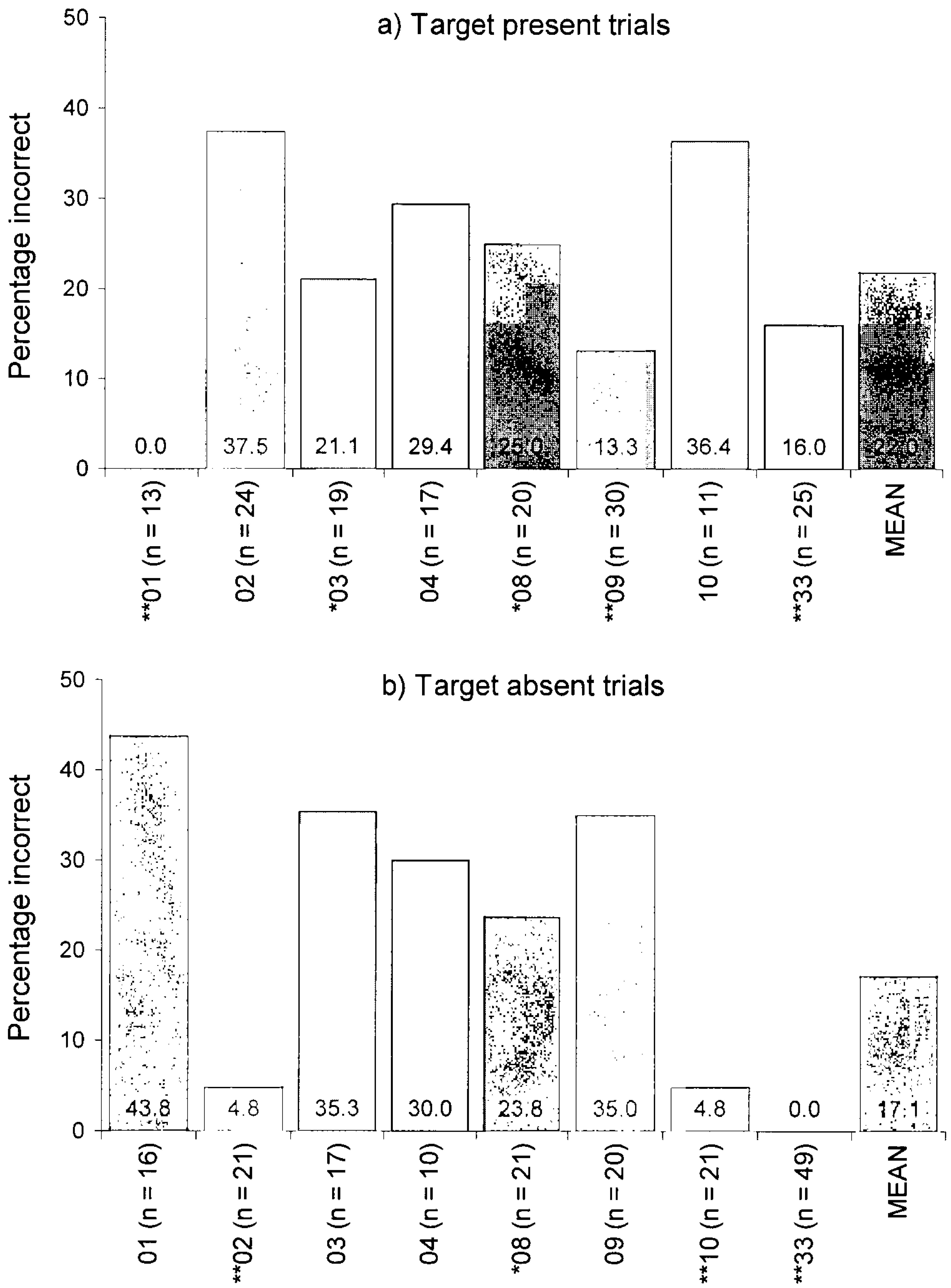



*Figure 6.2: Percentage error rate associated with each actor in a) target present and b) target absent trials. The letter indicates the actor physically present in each trial. The number of participants attending each session is provided in parentheses. Note: Results of Binomial test indicting individual actor identification sessions in which participants were significantly more accurate than would be expected by chance (* < .05; ** p < .001)*

### 6. 1. 2. 4. Confidence levels

To analyse converted confidence levels, data from individual actor sessions were pooled. The mean confidence for correct and incorrect decisions is reported in Table 6.1 as a function of participant decision ('same' or 'different').

*Table 6.1: Mean correct and incorrect adjusted confidence level scores as a function of decision type (standard deviations in parentheses)*

| | Same Decision | | Different Decision | |
|---|---|---|---|---|
| | Correct<br>$n = 124$ | Incorrect<br>$n = 30$ | Correct<br>$n = 145$ | Incorrect<br>$n = 35$ |
| Converted scores | 2.75<br>(0.91) | 1.77<br>(0.77) | 3.08<br>(0.98) | 2.23<br>(1.11) |

A 2 (accuracy) x 2 (decision) ANOVA revealed a significant main effect of accuracy, $F(1, 330) = 47.81, p < .001$; correct decisions were associated with higher confidence levels than incorrect decisions ($M = 2.93$, $SD = 0.96$ vs. $M = 2.02$, $SD = 0.99$). The effect of decision type was also significant, $F(1, 330) = 8.85, p < .005$; confidence was higher for 'different' decisions than for 'same' decisions ($M = 2.91$, $SD = 1.06$ vs. $M = 2.56$, $SD = 0.96$). The interaction was not significant, $F < 1$.

### 6. 1. 3. Discussion

Experiment 6.1 demonstrated that even when an unfamiliar person is physically present, identity verification from a video image is prone to error. In target present conditions, 22% of matching judgements were incorrect. These results are comparable to previous research using photographs as the target medium (e.g., Burton et al., 1999; Bruce et al., 2001). However, the results of some of the individual identification sessions highlight concerns as accuracy varied considerably. Performance in target present trials involving one actor was at ceiling, whereas it was at chance levels in trials involving three other actors. Bruce et al. (1999) found that even if photographs and videos were captured on the same day, errors at this task occurred. In Experiment 6.1, filming sessions took place three weeks prior to the identification sessions and it is possible that the appearance of the actors may have altered. Nevertheless, court proceedings would take place some time after an incident and therefore this experiment possessed ecological validity.

In target absent conditions, approximately 17% of participants wrongly identified a distracter present in person as being in the video. Again, performance, as measured using both the scale and accuracy data varied. In sessions involving four of the eight actors, error rates were over 30% and performance did not significantly differ from chance. Indeed, over 40% of participants wrongly identified one actor as being present in video. Previous research has found that unfamiliar face identification judgments are based mainly on external features (e.g., Bruce et al., 1999; Young et al., 1985), especially hairstyle (e.g., O'Donnell & Bruce, 2001), as this is probably the most salient cue. The high levels of misidentifications in this experiment may partly be due to the homogeneous appearance of the actors and a tendency for similar hairstyles. However, this group of young men might plausibly be the subject of a disputed identification. For example, a case may involve a number of people who admit presence at or near a crime scene, while denying they are the individual recorded by CCTV committing a crime (e.g., an assault outside a nightclub).

Confidence in responses in Experiment 6.1 was consistent with the results found in experiments reported in Chapters 4 and 5 as well as those reported by Bruce et al. (1999) and Henderson et al. (2001). Correct decisions were consistently associated with higher confidence than incorrect decisions. Confidence in 'different' decisions, or those indicating a belief that the live actor was not depicted in video was also consistently higher than for 'same' decisions. If replicated in a court room by jurors, this suggests that they would more cautious in implicating 'guilt' on the basis of video footage alone.

One potential issue of concern is that participant numbers were low in some cells and high percentage error rates in some cells may be a consequence of this. Furthermore, the majority of participants were undergraduate students whose responses have been found in many domains to not correspond with those of members of the general public (Foot & Sanford, 2004). In addition, Experiment 6.1 took place less than a month after the video images were taken, whereas criminal investigations can take considerably longer before a court appearance, even if a suspect is quickly apprehended. Finally, it is possible that the differences in performance across actors may partly be due to whether their gait was particularly distinctive. In a court room a guilty defendant might attempt to disguise any idiosyncratic movements captured on CCTV.

Therefore, Experiment 6.2 was designed to examine whether similar error rates would be found if firstly, participants from a wider demographic background were recruited; secondly, with an increase in the numbers of participants viewing each actor; thirdly, with a longer time interval between identification session and video footage capture and fourthly, with no provision of gait information as the actors did not walk across the room during the identification sessions. Finally, examination was also conducted to see whether the disguise effects found in Chapter 5 using photographs as targets would be replicated with live actors.

**Experiment: 6. 2. Live actor matching with footage one-year old**

Many criminal investigations can take months or even years to reach the courts. It is likely that in the intervening period the appearance of the defendant will have changed. In court, jurors would be aware of this and even if a defendant does not appear to strongly resemble the perpetrator shown in video footage, it is possible that they might unquestionably accept that it is the same person. The primary aim of Experiment 6.2 was therefore to examine whether knowledge that surveillance footage is not of recent origin would influence perceptual matching judgements. Half the participants were correctly warned that the videos were a year old. If they based their judgements on this age information, it was predicted to result in a bias towards responding that the actor was shown in video. This was expected to result in higher unadjusted identity-decision scale scores in both target present and target absent conditions leading to an increase in target absent errors, with a consequent reduction in target present errors.

A secondary aim was to examine whether the effect of disguise found using photographs in Experiment 5.1 would also occur when the targets were live in person, instead of in photographs. Although there were more potential identification cues available in this task when the actors were live, the face is the most important identifying feature (Burton et al., 1999). Therefore, the results were expected to be consistent with Experiment 5.1 and it was predicted that the wearing of disguise in video would increase the number of errors in target present conditions, especially when actors were wearing a hat. In contrast, disguise was not predicted to have any additional effects on performance in target absent conditions.

### 6. 2. 1. Method

#### 6. 2. 1. 1. Participants

Five hundred and ninety-one (303 male; 288 female) adult participants contributed data to this study (Mean age = 35.1, *SD* = 11.6). All were recruited at the Science Museum, London using the procedures described in Chapter 4. None had taken part in any of the previous experiments or pilot studies.

#### 6. 2. 1. 2. Actors and materials

Two of the matched actors employed in Experiment 6.1 were again recruited for this study. This meant that videos of Actor A03 were employed in target absent conditions involving Actor A04, and vice versa. Photographs taken at the time of obtaining the videos, and also at the time of the live identification sessions a year later are shown in Figure 6.3. The videos were those showing the actors in no disguise, in glasses or in a hat as described in Section 3.2.2.

#### 6. 2. 1. 3. Design

This experiment utilised a 2 (target presence) x 2 (warning) x 3 (disguise) x 2 (actor) independent measures design with a single-item identity-verification task. Participants were shown a video clip and had to decide using the identity-decision scale whether a physically present actor was depicted in the footage. The first factor, target presence, was whether the actor was present or absent in the video. The second factor, warning, was whether participants were warned in advance that the video was filmed one year ago or whether no warning was given. The third factor was the disguise of the 'culprit' shown in the video: no disguise; dark glasses or hat. As two different live actors were recruited, 'actor' was treated as a fourth variable. The dependent variables were unadjusted scores on the 8-point identity scale as well as accuracy and confidence measures obtained from score conversion.

*Figure 6.3: Photographs of Actors 03 (left) and 04 taken at the time of the original video filming sessions in February 2003; and posing together at the time of the Experiment 6.2 identification sessions in February 2004*

### 6. 2. 1. 4. Procedure

Participants, in groups of up to eight, were seated approximately 2m in front of a 16" television screen, with one of the two actors standing with arms folded next to the screen. The video footage sometimes showed the 'culprit' in one of three disguises and the 'live' actors present in the room stood still throughout the procedure. In all conditions, the actor in the room did not wear any 'disguise'. Approximately half the participants were verbally informed in advance that the films were a year old. The rest were given no additional information. A photograph depicting the viewing conditions in a target present trial involving Actor 03 is shown in Figure 6.4. Participants were closely watched to ensure no collaboration and recorded their responses using the eight-point identity-belief scale. Full performance feedback was provided.

*Figure 6.4: Viewing conditions in Experiment 6.2*

### 6. 2. 2. Results

#### 6. 2. 2. 1. Unadjusted scale data

Mean unadjusted identity-decision scale scores as a function of disguise, warning and target presence are reported in Table 6.2. A 3 (disguise) x 2 (warning) x (live actor) x 2 (target presence) independent measures ANOVA conducted on these data found that the main effects of disguise, $F < 1$; of live actor, $F < 1$; and of warning, $F(1, 567) = 1.36$, $p > 0.05$ were all non-significant. However, the main effect of target presence was highly significant, $F(1, 567) = 39.08$, $p < 0.001$; scores were higher when targets were present ($M = 4.71$; $SD = 2.48$) than when they were absent ($M = 3.45$; $SD = 2.33$).

The interaction between warning and disguise was also significant, $F(2, 567) = 3.06$, $p < 0.05$. Three Bonferonni-corrected independent measures t-tests were conducted to examine the simple effects of warning within each disguise condition. The warning was found to marginally increase scale scores in the dark glasses condition only, $t(190) = 2.00$, $p < .1$. There was no effect of warning on the other disguise conditions ($p > .1$).

*Table 6.2: Unadjusted scale data as a function of target presence, disguise and warning (standard deviations in parentheses)*

| | | No disguise | | Dark glasses | | Hat | |
|---|---|---|---|---|---|---|---|
| | | Warning | No warning | Warning | No warning | Warning | No warning |
| Target present | *n* | *50* | *46* | *45* | *54* | *45* | *53* |
| | Unadjusted scale scores | 4.42 (2.63) | 4.54 (2.67) | 5.93 (2.06) | 4.67 (2.26) | 4.38 (2.60) | 4.42 (2.40) |
| Target absent | *n* | *43* | *61* | *50* | *43* | *50* | *51* |
| | Unadjusted scale scores | 3.81 (2.38) | 3.66 (2.64) | 3.36 (2.17) | 2.91 (2.01) | 3.36 (2.35) | 3.55 (2.27) |

There was also a significant interaction between target presence and disguise, $F(2, 567) = 4.87$, $p < 0.01$. Two Bonferonni-corrected independent measures ANOVAs conducted to examine the simple effects of disguise found that there was a difference in scores in target present conditions, $F(2, 290) = 3.53$, $p < .05$. Simple comparisons found that scale scores were higher when actors were depicted in dark glasses than in the other two disguise conditions ($p < .05$), which did not differ ($p > .1$). There were no differences in scores in target absent conditions, $F(2, 295) = 1.48$, $p > .1$.

Finally the three-way interaction between warning, target presence and live actor was significant, $F(1, 567) = 3.92$, $p < 0.05$. Bonferonni simple interaction analyses on these data examined the interaction between target presence and warning for each actor. For Actor 03, the main effect of target presence was significant, $F(1, 263) = 22.62$, $p < .01$; target present scores were consistently higher than target absent scores. The effect of warning and the interaction were not significant ($p > .1$). The effect of target presence was also significant for Actor 04, $F(1, 320) = 17.84$, $p < .001$; with scores again higher in target present conditions. The main effect of warning was not significant (p > .1); however, the interaction between target presence and warning approached significance, $F(1, 320) = 3.29$, $p < .1$; simple comparisons found that scores were higher when a warning was given in target present conditions only ($p < .05$). A warning had no effect in target absent conditions ($p > .1$). No other interactions were significant ($p > .1$).

### 6. 2. 2. 2. Accuracy data

Table 6.3 shows the adjusted percentage error rate as a function of disguise, warning and target presence. Across all trials, 38.6% of responses were incorrect. These data were analysed as a function of the proportion of errors in each condition. A series of 12 Binomial tests conducted on each cell in Table 6.3 to examine whether performance was significantly better than chance, are also reported within this table. Performance was found to be significantly more accurate than chance in only one target present condition, when a warning was given concerning the age of the footage and the actor was wearing dark glasses. In contrast, in target absent conditions, performance was at chance levels when the actors were depicted wearing no disguise only, regardless of whether a warning was given or not.

*Table 6.3: Percentage error rate as a function of target presence, disguise and warning*

| | | No disguise | | Dark glasses | | Hat | |
|---|---|---|---|---|---|---|---|
| | | Warning | No warning | Warning | No warning | Warning | No warning |
| Target present | Percentage error rate | 50.0 | 47.8 | 20.0** | 40.7 | 48.9 | 54.7 |
| Target absent | Percentage error rate | 39.5 | 39.3 | 34.0* | 23.3** | 28.0* | 33.3* |

*Note: Results of Binomial test indicting individual identification sessions in which participants were significantly more accurate than expected by chance (* < .05; ** p < .001)*

A 3 (disguise) x 2 (target presence) x 2 (warning) x 2 (actor) independent measures ANOVA on the accuracy data revealed a significant main effect of disguise, *F*(2,

567) = 5.24, $p$ < .01; Bonferonni post hoc tests found that the fewest errors were made in the glasses condition (30.2%), significantly less than in the no disguise condition (44.0%, p < .01) and marginally fewer than in the hat condition (41.2%, $p$ = 0.068). The latter two conditions did not significantly differ ($p$ > .1). The effect of target presence was also significant, $F$(1, 567) = 7.69, $p$ < .01; more errors were made when targets were present (44.0%; $SD$ = 49.7) than when they were absent (33.2%; $SD$ = 47.2). The main effects of warning and of actor were not significant, $F$ < 1.

However, the interaction between warning and live actor was significant, $F$(1, 567) = 7.03, $p$ < .01 (Figure 6.5). Bonferonni-corrected simple effects found that when Actor 04 was live, the error rate was lower when a warning was given, $t$(322) = 2.48, $p$ < .05. In contrast, when Actor 03 was live this effect was reversed, although the difference was not significant ($p$ > .1).

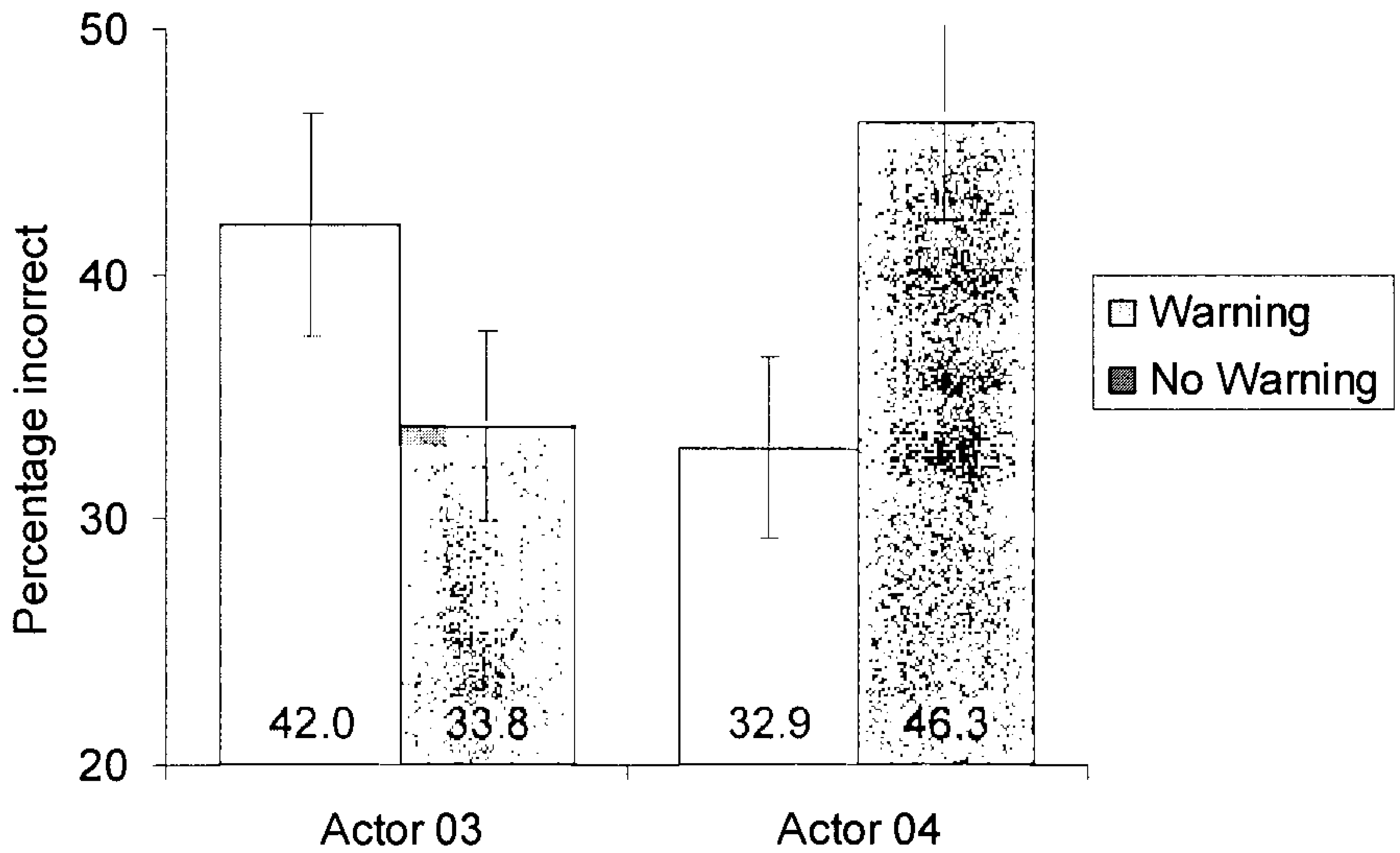



*Figure 6.5: Mean percentage error rates as a function of live actor and warning (Error bars indicate standard error of the mean)*

The three-way interaction between disguise, target presence and warning also approached significance, $F$(2, 567) = 2.38, $p$ = .093 (Figure 6.6). Bonferonni-corrected simple interaction effects analyses found that when targets were in no disguise, the effects of target presence and warning and the two-way interaction were not significant ($p > .1$).

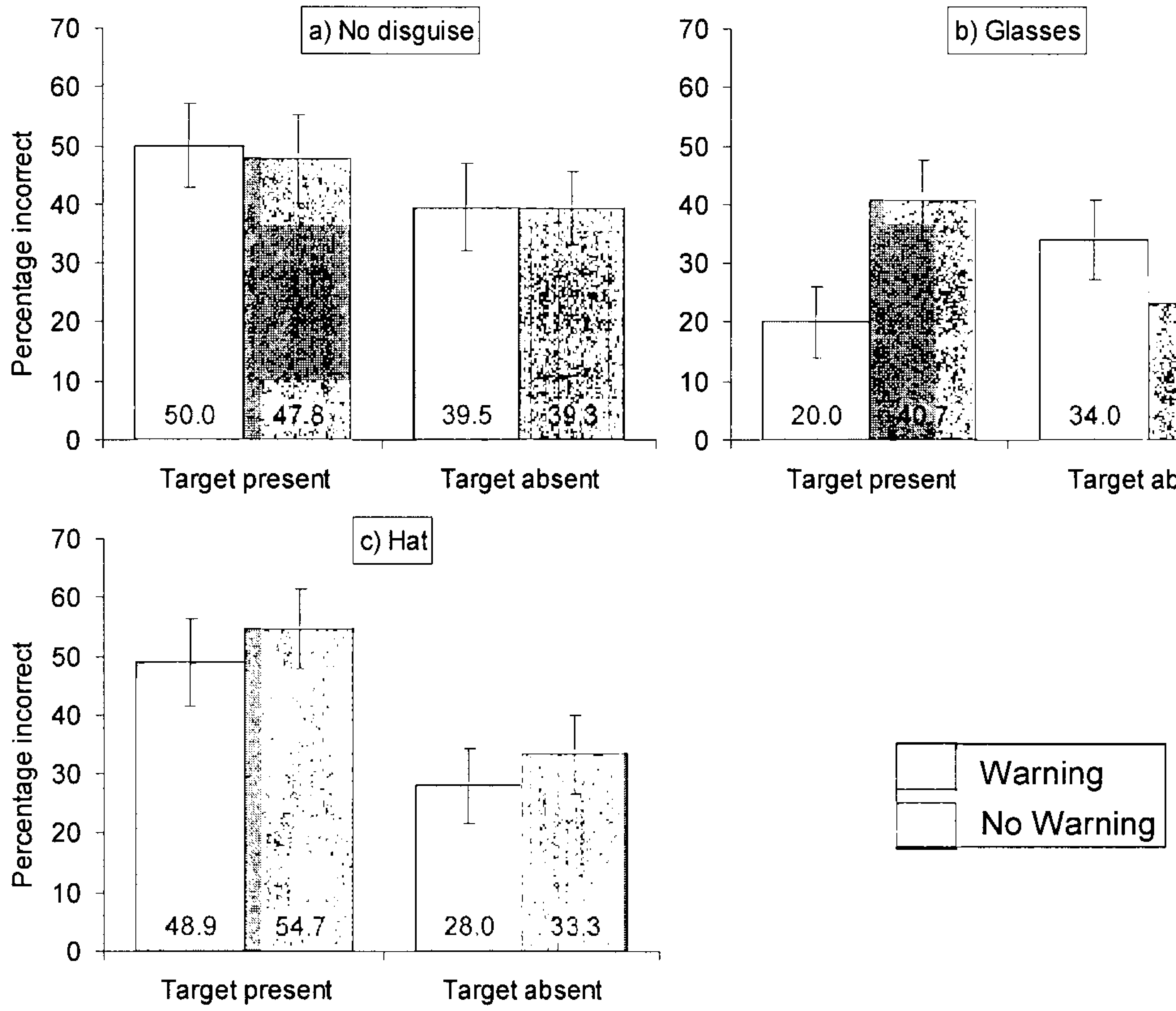



*Figure: 6.6. Percentage error rates as a function of target presence and warning when actors were depicted in a) no disguise; b) dark glasses and c) a hat (error bars denote standard error of the mean)*

When the actors wore dark glasses on video, the effects of target presence and warning were also not significant ($p > .1$). However, the interaction between actor and warning was significant, $F(1, 188) = 5.67, p < .05$. Simple comparisons found a warning reduced errors by more than a half when targets were present although the effect was only marginally significant, $t(97) = 2.25, p < .1$. In contrast, a warning increased errors one-and-a-half times in the target absent condition although this was not significant ($p > .1$).

Finally, when targets were depicted in a hat there was a significant effect of target presence only, $F(1, 195) = 9.43, p < .01$, more errors were made when targets were present. The effect of warning and the interaction were not significant ($p > .1$). No other interactions were significant ($p > .1$).

#### 6. 2. 2. 3. Confidence level data

To examine converted confidence levels, data from the two actors were pooled to increase statistical power. A 2 (accuracy) x 3 (disguise) x 2 (decision) x 2 (warning) independent measures ANOVA on these data found a significant main effect of accuracy, $F(1, 567) = 8.81$, $p < .005$; correct responses were associated with higher confidence than incorrect responses ($M = 2.90$, $SD = 0.97$ vs. $M = 2.68$, $SD = 1.00$). The main effect of disguise was also significant, $F(2, 567) = 3.96$, $p < .05$; Bonferonni post hoc tests found that levels were higher in the no disguise condition than when actors were wearing glasses ($M = 2.96$, $SD = 0.94$ vs. $M = 2.71$, $SD = 0.99$; $p < .05$). Confidence levels in the hat condition did not differ significantly from the other two disguises ($M = 2.78$, $SD = 1.02$; $p > .1$). The main effect of decision was also significant, $F(1, 567) = 18.44$, $p < .001$; confidence in 'different' decisions was higher than in 'same' decisions ($M = 2.97$, $SD = 0.97$ vs. $M = 2.63$, $SD = 0.97$). However, the main effect of warning, $F < 1$ and all interactions were non-significant ($p > .1$).

### 6. 2. 3. Discussion

Experiment 6.2 confirmed the difficulties inherent in simultaneous matching from video to live actors. The footage had been taken one year previously and when targets were absent, 33% of participants were incorrect by responding that the actor physically present was also depicted in the video. However, more errors were made in target present conditions (44%), possibly due to the one-year time interval between video capture and identification session.

Indeed, in Experiment 6.2, 47.8% of responses were in error in target present conditions when no warning was given and the actors were in no disguise. These results can be directly compared to those of Experiment 6.1 in which the same two matched actors, also wearing no disguise were employed (Actors 03 and 04). Experiment 6.1 took place only three weeks after the videos were taken and the error rate was almost half that of Experiment 6.2 (25.0%). To some extent, this difference must reflect alterations to the actor's appearance over the 11 month time period. However, in target absent conditions, the error rates to the same two actors were approximately 33.3% in Experiment 6.1 but again were substantially higher at 39.3% in Experiment 6.2. This also suggests that the removal of gait cues in Experiment 6.2 made it harder for participants, both to differentiate between the two actors, and to determine whether the same actor was shown in video in target present conditions.

Providing a warning as to the age of the footage did not have the predicted effects on the accuracy of responses. It had no effect when the actors were depicted in no disguise or a hat. Instead, a warning introduced a bias when actors wore dark glasses only. This effect was subtle, reducing the number of target present errors while slightly increasing target absent errors. Indeed, when targets were present, performance was at chance levels in all conditions except when the warning was given and the actors were depicted in dark glasses. Research has found that typical faces are recognised less effectively than distinctive faces, reducing hit rates and increasing the number of false alarms (e.g., Valentine, 1991). The wearing of glasses has been found to make faces appear more homogenous, thus reducing distinctiveness (e.g., Terry, 1993; 1994). The results of this experiment are consistent with an increase in the perceived typicality of the faces shown in video when wearing dark glasses. This homogeneity effect may have meant that participants were less conservative in their identification decisions when warned of the age of footage, with the subsequent impact on error rates in this condition only.

The warning also had the effect of improving accuracy when Actor 04 was live, regardless of whether he was in the video or not and the reverse though non-significant effect of reducing accuracy in sessions involving Actor 03. Reasons for these effects are unclear as the improvement in performance for Actor 04 involved the presentation of six videos, three of each actor in the disguise conditions. It would suggest that knowing the video was taken a year previously allowed participants to utilise specific features likely to be stable over time in the videos, and compare them specifically to Actor 04 in order to make their choices more accurate. However, these data support the contention of Wright and Sladden (2003), in that caution should be taken if generalising results obtained from single actor studies to the wider population, as effects can be specific to individuals.

The overall effect of disguise was also not as predicted. Errors were highest when targets were in no disguise and lowest when they were wearing glasses. These results are not consistent with those found in previous studies which have found that disguise reduces matching and recognition accuracy in photographs (e.g., Diamond & Carey, 1977; Flin, 1985a; Henderson et al., 2001, Hockley et al., 1999; Metzger, 2001; Patterson & Baddeley, 1977; Terry, 1994). Furthermore, in Experiment 5.1 and 5.2 in which the actors in the videos were matched to actors depicted in photographs, disguise reduced matching accuracy, but only when targets were present. These contradictory results therefore suggest that disguise may have less influence on identity decisions when actors are live in person, possibly due to the increased availability of more cues than are available in photographs. However, only two actors were employed in Experiment 6.2, whereas images of 24 actors were included in the experiments reported in Chapter 5. It is possible therefore that these contradictory disguise results in Experiment 6.2 might not generalise to the wider population.

**Experiment: 6. 3. Live actor matching by adults and children**

Two of the experiments reported in this thesis have compared the face matching ability of adults of different ages and of children (Experiments 4.1 & 4.2). In Experiment 4.1, children under the age of 10 were found to be inferior to adults in matching people depicted on video to facial photographs in arrays. Furthermore, adolescent children aged 12 – 14, and older adults were also found to be moderately worse at the task than young adults. These results were consistent with those found in previous face matching and recognition studies (e.g., Carey et al., 1980; Smith & Winograd, 1978).

However, no published matching study appears to have been conducted using children as participants when the target is present in person. In an eyewitness experiment, assessing memory, Dent (1977) found that children were less likely to correctly select a live target than a photograph. However, there were no differences in the proportion of incorrect selections from the lineups. She suggests that these results were due to embarrassment and stress meaning that the children were less likely to spend time inspecting the actors, preferring to make a 'not present' selection. In contrast, they would closely examine photographs before making a selection. In a later study, she found that reducing stress levels in children reduced the difference between live and photographic lineups (Dent & Stephenson, 1979). In addition, a study by Peters (1991) found that if children were introduced to an adult given money by a teacher in a 'low stress' condition, selection from a deliberately 'low-stress' live lineup was equal to that when photographic lineups were shown. In contrast, when a stranger walked into the room and stole money in a 'high-stress' condition, children were less likely to identify the target in a 'high-stress' live lineup, some being upset and telling their parents later that they did not want to get the 'robber' into trouble. Performance on photographic spreads was equal, regardless of whether children had seen the high or low stress encounters.

In a previous study by Dent & Gray (1975, cited in Dent, 1977), adult participants showed a similar pattern in a stress-based design, with an advantage for photograph lineups over live lineups. In contrast, other recognition studies examining adults only have found null effects (Shepherd et al., 1982), or an advantage for live line ups over photographs (e.g., E. Brown et al., 1977; Egan et al., 1977). The studies reported above all used lineup designs or arrays of distracters as these are common practice when witnesses are required to identify a suspect. However, in some circumstances identification decisions will be made to single suspects, for instance, when making a street identification, or 'show-up'. The single-item identity-verification design in Experiment 4.2 is an analogue of that scenario as although the memory of participants was not tested; matching judgements were made to single actors shown on video and in photographs. In that experiment no differences in accuracy across age groups were found, although the youngest children tested (8 – 11 year-olds) were more confident in their decisions, regardless of accuracy or whether they believed that both images depicted the same person or not.

Children had originally participated in Experiment 6.2. However, the majority only viewed one of the two actors (Actor 04). Therefore, to avoid confounding the results, their data was omitted from that experiment. These results are reported here, along with those from adults who had viewed the same actor. Results from both adult and child participants who had viewed Actor 03 are omitted. Effects that were consistent with those found in Experiment 6.2 are not further discussed. Of interest are any differences found between children and adults.

However, due to the age profile of the children recruited for Experiment 6.3, it was not possible to separate them into the same discrete age categories as had occurred in Experiments 4.1 and 4.2. Therefore participants in Experiment 6.3 were split into two age groups only, those aged 18 and above (adults), and those aged 17 and under (children). Nevertheless, experimental predictions based on the findings of Experiment 4.2 could be made, as the results of trials involving Actor 04 in Experiment 4.2 were consistently in line with the overall results of that experiment. In addition Actor 04 was matched with Actor 03 in target absent trials in both Experiments 4.2 and 6.3. Furthermore, the results of Experiment 4.2 can be broken down into the same two discrete age group categories that were used in Experiment 6.3.

As such, in Experiment 4.2, there were no differences in accuracy between adults and children, in terms of trials involving Actor 04 alone, or in terms of the overall results of the experiment. However, children were found to be more confident than adults in their decision making. Therefore, as the single-item identity-verification design for this experiment was similar to that when matching to photographs in Experiment 4.2, it was predicted that the results would be consistent. As such, no age group differences were expected when measuring accuracy. However, children were expected to demonstrate higher decision confidence.

### 6. 3. 1. Method

#### 6. 3. 1. 1. Participants

One hundred and twenty-seven children (61 male; 66 female) under the age of 18 years contributed data to this experiment (Mean age = 11.1, *SD* = 2.5). These data were combined with the adult data from the same identification sessions involving Actor 04 only ($n$ = 324; 169 male, 155 female) described in Experiment 6.2 (Mean age = 34.9, *SD* = 11.5). The data from sessions involving Actor 03 were omitted from all analyses (adults: $n$ = 267, children: $n$ = 8). None of the participants had taken part in any of the previous experiments or pilot studies.

#### 6. 3. 1. 2. Design and procedure

The experiment employed a 2 (age group) x 2 (warning) x 3 (disguise) independent-measures single-actor identity-verification design. Conditions were as described in Experiment 6.2.

### 6. 3. 2. Results

Table 6.4 displays the mean unadjusted scale data and percentage error rates as a function of age group, target presence, disguise and warning.

#### 6. 3. 2. 1. Unadjusted scale scores

The first analysis compared adults and children on raw scale data as a function of target presence, warning, and disguise, as this provides indications of any response bias. A 2 (age group) x 2 (target presence) x 2 (warning) x 3 (disguise) independent-measures ANOVA revealed a significant main effect of age group,

$F(1, 427) = 11.91$, $p < .001$, children's scores were higher than adults ($M = 5.16$, $SD = 2.55$ vs. $M = 4.23$, $SD = 2.54$). The effect of target presence was also significant, $F(1, 427) = 10.39$; $p < .001$; scores were higher in target present than in target absent conditions ($M = 4.98$, $SD = 2.57$ vs. $M = 3.99$, $SD = 2.48$). There was a marginal effect of warning $F(1, 427) = 3.12$, $p = 0.078$; scores were slightly higher when a warning about the age of the footage was given than when no information was provided ($M = 4.65$, SD = 2.59 vs. $M = 4.32$, $SD = 2.55$). The effect of disguise was not significant, $F(2, 427) = 2.18$, $p > .05$.

*Table 6.4: Unadjusted scale scores (standard deviations in parentheses) and percentage error rates as a function of age group, target presence, disguise and warning*

| | Target present | | | | | | Target absent | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | No disguise | | Glasses | | Hat | | No disguise | | Glasses | | Hat | |
| | W | NW | W | NW | W | NW | W | NW | W | NW | W | NW |
| | Adults | | | | | | | | | | | |
| Mean Scale scores | 5.23 (2.56) | 4.42 (2.70) | 6.00 (2.18) | 4.70 (2.23) | 4.20 (2.86) | 4.21 (2.58) | 4.05 (2.52) | 4.11 (2.66) | 3.09 (2.16) | 3.39 (2.31) | 3.59 (2.37) | 3.85 (2.43) |
| Percentage error rate | 33.3 | 50.0 | 16.0 | 40.0 | 52.0 | 62.1 | 45.5 | 46.4 | 30.3 | 34.8 | 24.1 | 42.3 |
| | Children | | | | | | | | | | | |
| Mean Scale scores | 6.45 (2.16) | 4.44 (3.32) | 6.58 (1.51) | 4.45 (2.77) | 5.90 (2.51) | 4.64 (2.50) | 5.20 (2.53) | 6.00 (1.76) | 5.17 (2.48) | 4.55 (3.11) | 3.50 (2.22) | 4.70 (2.67) |
| Percentage error rate | 18.2 | 44.4 | 8.3 | 45.4 | 30.0 | 54.5 | 70.0 | 70.0 | 66.7 | 45.4 | 30.0 | 60.0 |

*Note: W = warning; NW = no warning.*

There was a significant interaction between warning and target presence, $F(1, 427) = 9.31$, $p < .005$ (illustrated in Figure 6.7). Simple effects analyses found that a warning significantly increased scale scores in target present conditions, $t(225) = 3.07$, $p < .01$. However, a warning had no effect on target absent scores ($p > .1$). No other interactions were significant ($p > .1$).

From these findings it can be seen that the only significant effect differentiating adults and children, concerned the main effect of age group, indicating that with consistently higher scores there was a bias for children to more confidently respond that the actor was shown in the video, regardless of target presence, disguise or warning.

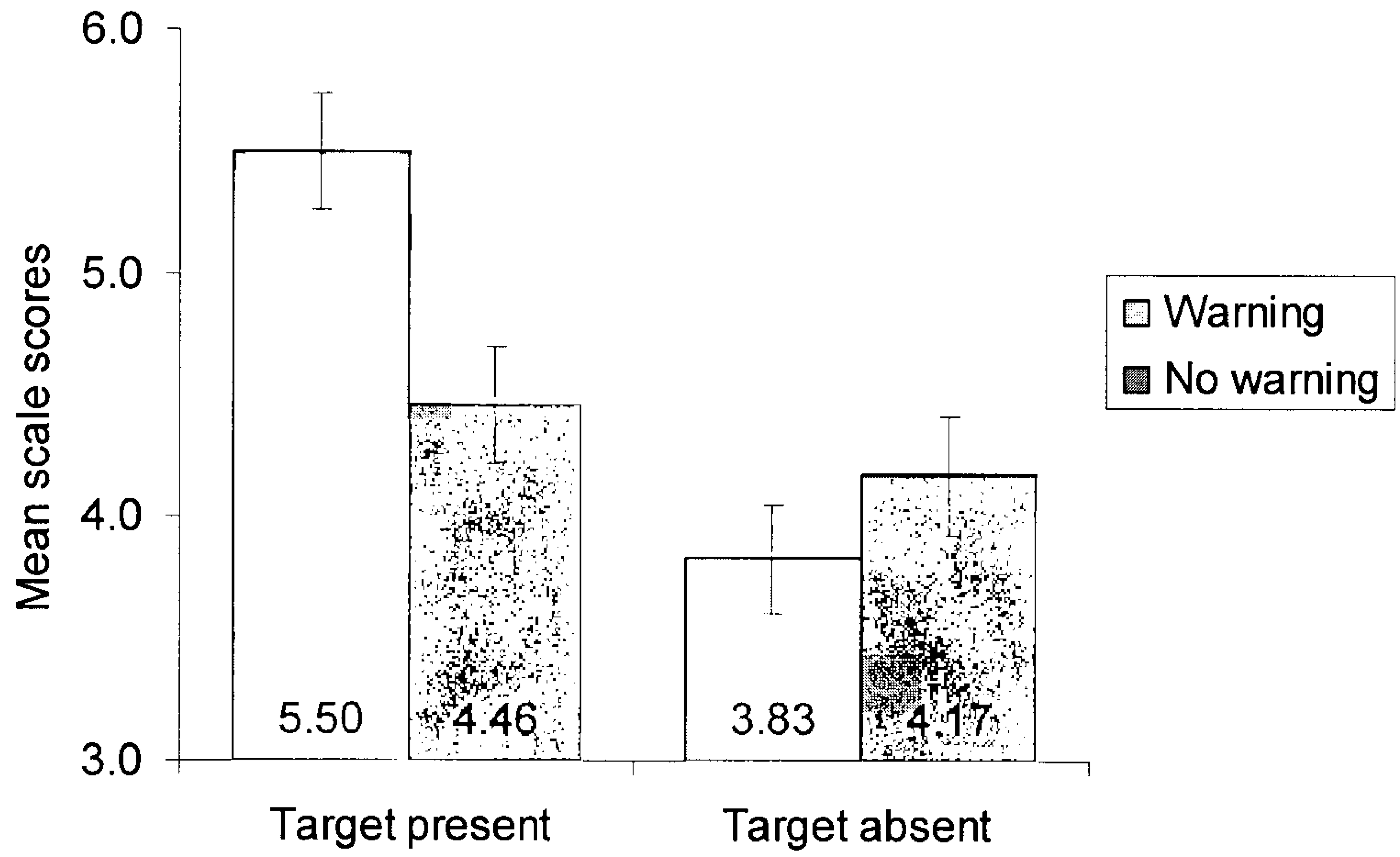



*Figure 6.7: Mean scale scores as a function of warning and target presence (error bars denote standard error of the mean)*

### 6. 3. 2. 2. Error rate data

Scale scores were converted for accuracy and these data are reported in Table 6.3. Overall, 41% of all responses were inaccurate, 39.7% ($SD$ = 49.0) in target present conditions, 42.4% ($SD$ = 49.5) in target absent conditions. A 2 (age group) x 3 (disguise) x 2 (target presence) x 2 (warning) ANOVA conducted on this data found that the main effects of age group, $F(1, 427) = 1.19$, $p > .05$; and of disguise, $F(2, 427) = 1.86$, $p > .05$; were non-significant. However the main effect of target presence was marginally significant, $F(1, 427) = 3.35$, $p = .068$; more errors were made in target absent conditions. The main effect of warning was also significant,

$F(1, 427) = 7.93, p < .005$; errors were higher when no warning was given (48.2%, $SD = 50.1$ vs. 34.1%, $SD = 47.5$).

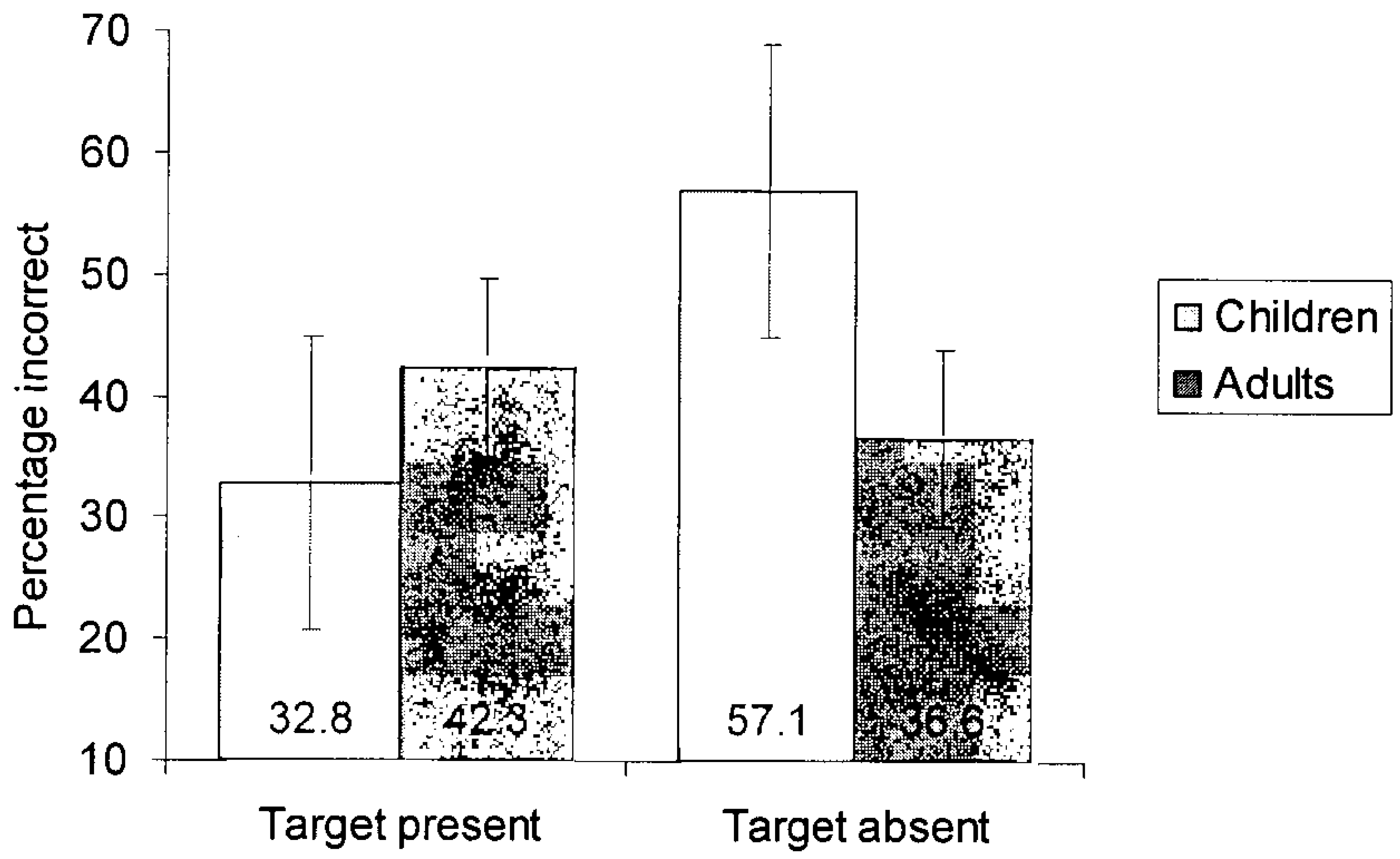



*Figure 6.8: Percentage error rates as a function of target presence and age group (error bars denote standard error of the mean)*

The interaction between disguise and target presence was significant, $F(2, 427) = 3.84, p < .05$. Simple effects analyses found no differences in target absent error rates across the disguise conditions ($p > .05$). However, there were differences across disguises when targets were present, $F(2, 224) = 5.29, p < .01$. Simple comparisons found that more errors were made in the hat condition than when the actor wore glasses (53.3%, $SD = 50.2$ vs. 28.2%, $SD = 45.3, p < .05$). However, the error rate for the no disguise condition (37.8%, $SD = 48.8$) did not significantly differ from the other two disguises ($p > .1$).

The interaction between target presence and age group was also significant, $F(1, 427) = 7.94, p < .005$ (Figure 6.8); simple effects analyses found that children made more errors than adults when targets were absent, $t(222) = 2.83, p < .05$. The reverse was found when targets were present although this difference was not significant ($p > .1$).

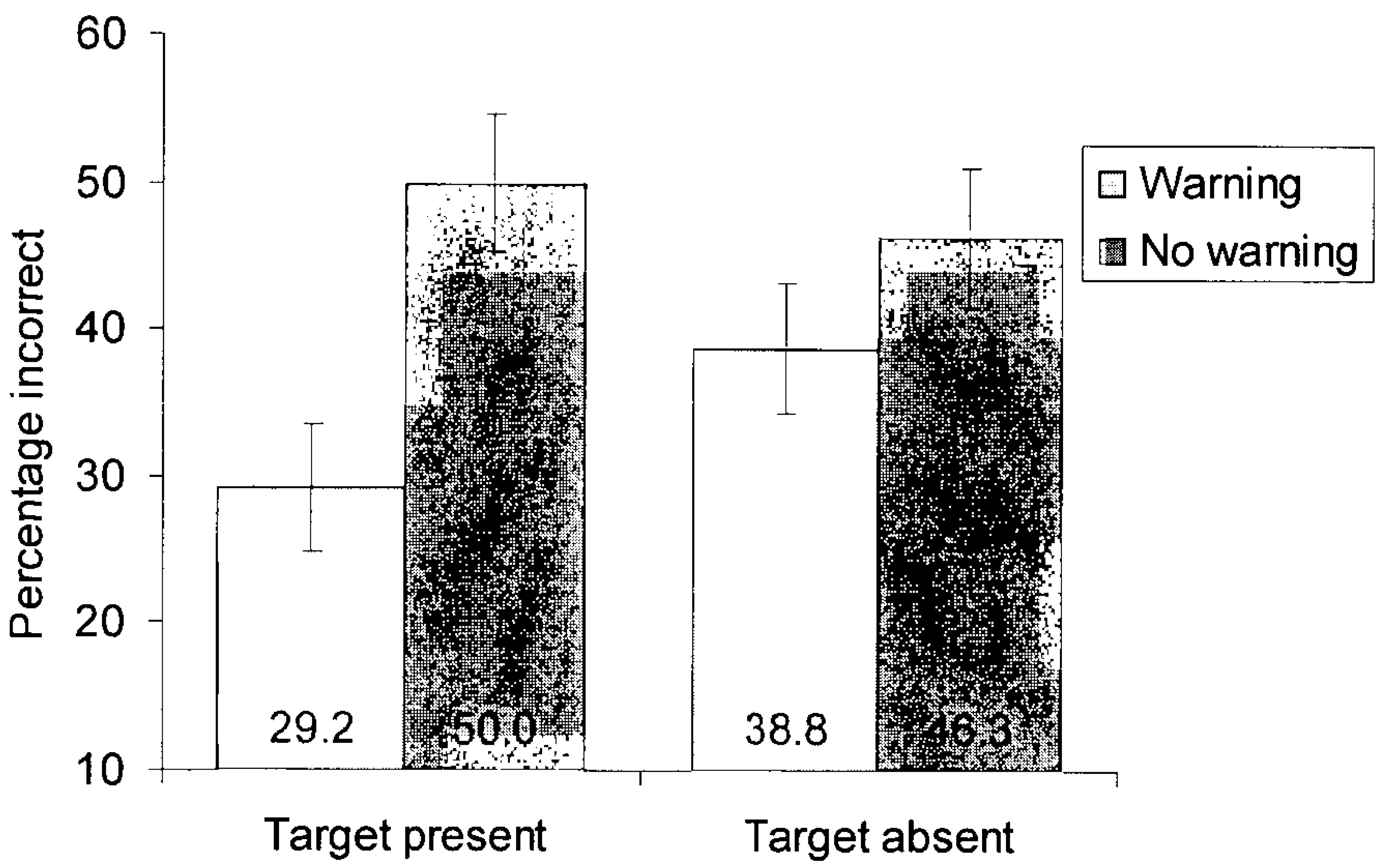



*Figure 6.9: Percentage error rates as a function of target presence and warning (error bars denote standard error of the mean)*

There was also a marginally significant interaction between warning and target presence, $F(1, 427) = 3.06$, $p = .081$ (Figure 6.9). Bonferonni simple effects analyses found that a warning reduced errors in target present trials, $t(223) = 3.26$, $p < .05$. However, a warning had a non-significant positive effect on accuracy in target absent trials ($p > .1$). None of the other interactions were significant ($p > .1$).

#### 6. 3. 2. 3. Confidence level data

Scale scores were also converted to provide confidence level data to examine age group confidence as a function of accuracy and decision type ('same'/'different'). A 2 (age group) x 2 (decision type) x 2 (accuracy) ANOVA conducted on these data found that the main effects of age group, $F < 1$; decision type, $F < 1$; and accuracy $F(1, 433) = 1.22$, $p > .1$, were not significant. However, there was a significant interaction between decision type and age group, $F(1, 443) = 4.47$, $p < .05$. Bonferonni-corrected simple effects analyses revealed that for adults, confidence was higher when making 'different' decisions than when making 'same' decisions, $t(322) = 2.72$, $p < .05$. For children, decision type had no effect on confidence levels and was the same as adults when making 'different' decisions ($p > .1$). This interaction is plotted in Figure 6.10. None of the other interactions was significant ($p > .1$).

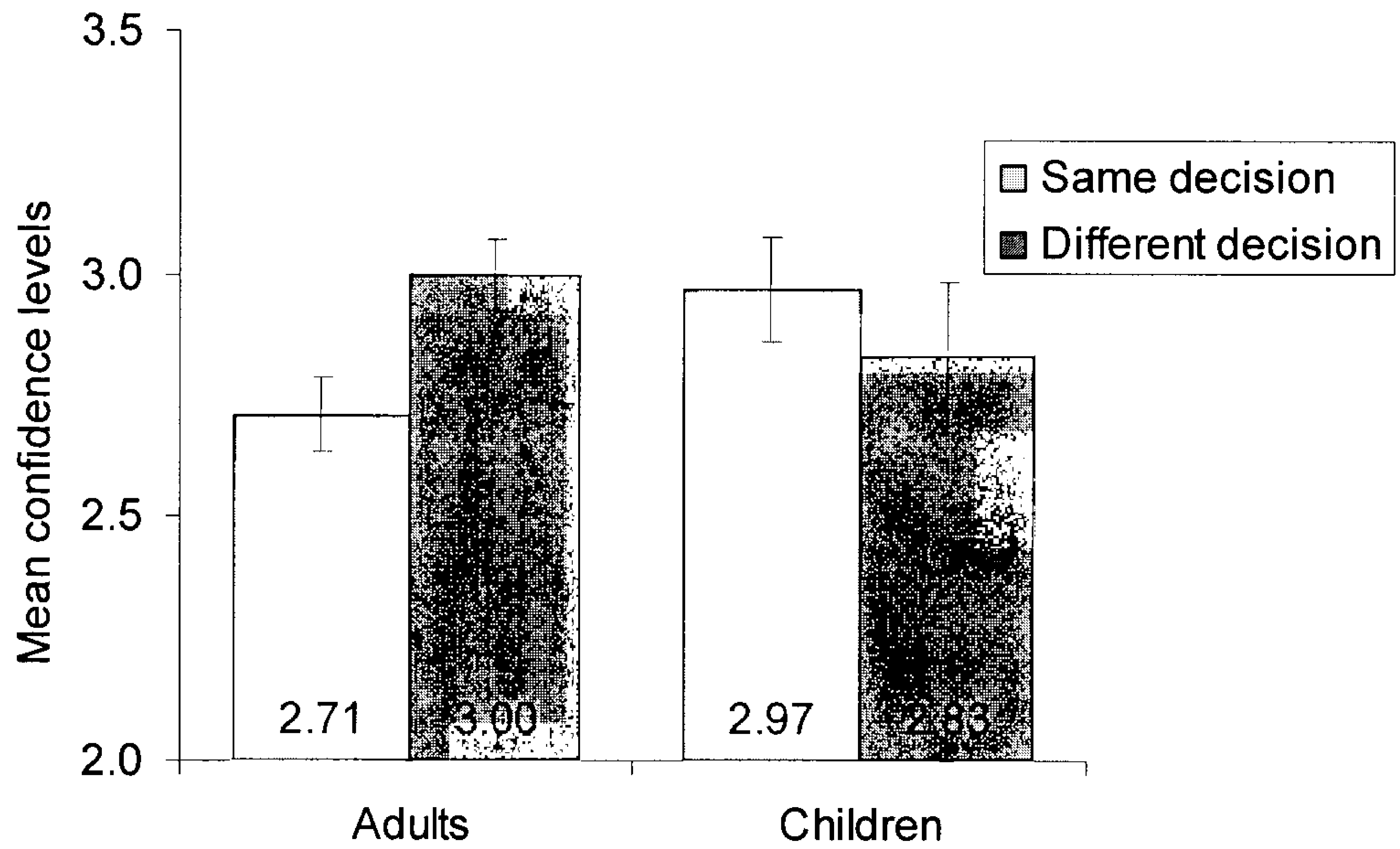



*Figure 6.10: Mean confidence levels as a function of age group and decision type (error bars denote standard error of the mean)*

**6. 3. 3. Discussion**

In Experiment 6.3, age group differences were found in the accuracy of participants to decide whether an actor present in person was depicted in video. Children's unadjusted scale scores were consistently higher than adults, pointing to a bias towards responding that the live actor was shown in the video regardless of whether this was correct or not. When converted into accuracy data, this liberal criterion resulted in more errors by children than adults in target absent trials and although not significant, fewer errors than adults in target present conditions. Furthermore, children were more confident when making 'same' decisions regardless of accuracy than adults. A similar propensity for children to select more targets than adults from lineups has been found in eyewitness identification experiments (Lindsay et al., 1997; J.F. Parker & Carranza, 1989; J.F. Parker & Ryan, 1993). J.F. Parker and colleagues suggested this may be due to a desire by children to please experimenters by responding positively, as well as to children believing that the experiment would not be meaningful unless the target is present.

However, these results contrast with the null effects of age group in terms of accuracy found in Experiment 4.2 in which a similar design was employed, only instead of a live actor, photographs were the target medium. There are a number of potentially confounding explanations for the different conclusions to these two experiments. Firstly, in Experiment 4.2 participants made decisions to six pairs of videos and photographs from a larger pool of actors, whereas decisions were made to a single live actor in Experiment 6.3 (Actor 04). However, analyses on the data using the target photograph of Actor 04 only in Experiment 4.2 was consistent with the results of the entire data set in that experiment, meaning that the effects could not be due to the specific video and photographic stimuli.

Furthermore, the results of other experiments in this thesis in which participants completed multiple trials were consistent with those in which they took part in a single trial; therefore, this explanation can be discounted. In addition, in some trials in Experiment 6.3, actors were depicted in disguise, therefore making the task more difficult. In contrast, in Experiment 4.2 all videos depicted actors in no disguise. Furthermore, in Experiment 6.3 approximately half the participants were correctly warned that the videos were a year old, whereas the photographs used in Experiment 4.2 were taken at the same time as the videos. However, disguise and warning did not interact with the age group variable on either the scale or accuracy data in Experiment 6.3 and these potential explanations can therefore also be discounted as inducing the age group differences.

However, there are similarities between the results of Experiment 6.3 and Experiment 4.1 in which children and adults made matching decisions to arrays. In both experiments, children were more likely to select the wrong actor. In Experiment 4.1 the majority of children (8 - 9 and 12 - 14 year-olds, but not 10 – 11 year-olds) were more likely to select an incorrect face from an array during a face matching task than the younger adults. In Experiment 6.3, children were more likely than adults to believe that an actor live in person was depicted in a video, even if someone else was present. It therefore appears that children are less able to distinguish between two images of different people, as well as being less able to accurately determine that a live actor is not present in an image, when a different actor is actually present. When the task is relatively simple, as in Experiment 4.2 in which participants viewed a series of single photographs and videos only, this inability does not impact on performance.

The results contrasted with those of previous studies (e.g., Dent, 1977; Dent & Stephenson, 1979; Peters, 1991) finding that children preferred to make 'not present' selections of live actors in line-ups in eyewitness identifications. Dent argued this was due to embarrassment and anxiety meaning that the children avoided viewing the actors. Although the results of Experiment 6.3 differ from the above, it may be that in this experiment children were again avoiding intensive inspection of the actors, making positive identifications based on a brief view of the resemblance between the actor and the video image.

### 6. 4. General discussion

The experiments reported in this chapter confirmed that when typical video footage obtained from open-street CCTV systems is available; unfamiliar face matching is prone to error even if the target is physically present. In published studies involving photographs, and in the previous experiments reported in this thesis, images from both mediums have been captured at approximately the same time and yet errors have still been found (e.g., Bruce et al, 1999; Henderson et al., 2001). In the two designs reported in this chapter, live identification sessions occurred a few weeks, and one-year after filming, with, as would be expected, an increase in errors associated with the longer time interval. Even though some participants were warned about this time interval in advance, accuracy only improved in trials associated with one of the two actors and in trials in which both actors were depicted in video wearing glasses.

Juries may be invited to conclude that a defendant resembles a perpetrator shown on video (*R v Blenkinsop*, 1995; *R v Dodson and Williams*, 1984; *R v McNamara*, 1996). This practice is supported in a review of the legislation (Attorney General's Reference, No. 2 of 2002, 2003). The experiments described in this chapter suggest that such an identification decision may be unreliable, especially as serious crimes often take months to reach court. Indeed, the age of footage warning bias found in Experiment 6.2, indicates that under some circumstances, people may be even more inclined to believe that two images of different people are of the same person.

The implications of the target absent errors are of most concern as these represent misidentifications and if replicated in court, would result in wrongful convictions. The actors were from a much smaller database than the 200 police recruits used by Bruce et al. (1999), and the several hundred actor-agency photographs examined by Henderson et al. (2001) to select their matched faces for arrays. This suggests that there may be many individuals in the general population who would be easily confusable in footage of this quality. It would not be inconceivable that these particular actors could be caught on CCTV wearing similar clothes (e.g., sports clothing). Therefore if any were accused of criminal activity, they might be primarily identified on this basis. However, Actor 03 (height: 1.72m; weight: 83kg) and Actor 04 (1.83m; 92kg) would actually be unlikely to be mistaken for one another for long during the course of a 'real' criminal investigation. If filmed by CCTV it might be possible, dependent on background details and geometry of camera angles, to calculate their approximate height and rule out the 'innocent' suspect. Nevertheless, the error rates to these actors were substantial, indicating that actors with an even closer resemblance would raise error rates further.

However, the video footage used in the experiments reported so far in this thesis was designed to simulate the quality that might be obtained from typical open-street systems. This raises the question of whether face matching would be more accurate if high-resolution images were available. Therefore, the experiment reported in the following chapter was specifically designed to examine face matching ability using high-quality close-up video footage.

# Chapter 7: Face matching with high-quality close-up video footage

## 7. 0. Introduction

The experiments reported in Chapters 4 to 6 demonstrated the difficulties inherent in identity matching from medium-range simulated CCTV footage. However, the face is the most important human identifying feature and in these images it may not have been clear enough for accurate performance. Nevertheless, previous published matching studies utilising much higher-quality close-up video footage have found similar results. For instance, Bruce et al. (1999) described a series of experiments using video footage showing head-and-shoulders shots of police recruits turning their faces from side-to-side. In these experiments, participants were required to make matching judgements to simultaneously presented arrays of photographed faces. In some, the footage was presented as moving videos; in others extracted stills were shown. The videos and photographs were taken on the same day.

In their second experiment the faces of the targets shown in a video still were always present in the arrays of ten photographs. Participants were aware of this, meaning the task was to select the face from the array that most resembled the target in video. Twenty-two per cent of trials involved the incorrect selection of a distracter when facial expression and pose matched. Higher error rates were found when viewpoint was different (32%). However, in one specific trial, 80% of participants failed to select a target, even when pose and expression was matched. Bruce et al. acknowledged that it would be unlikely that video footage of this quality would be obtained in normal criminal investigations. It would also be improbable that criminals during their activities would deliberately pose in this manner. Nevertheless, the authors describe the experiments using this footage as measuring the 'Gold Standard' of performance in this context.

However, in a courtroom a jury would be able to directly compare a single physically present defendant with CCTV footage. Therefore, the main aim of the experiment reported in this chapter was to replicate these circumstances using a forensically valid design with videos of similar quality to those used by Bruce et al. (1999).

**Experiment: 7.1: Face matching with close-up video images**

High-quality close-up video footage was obtained showing the actors' faces slowly turning $180^0$ in a continuous loop from left-to-right profile and back. This footage would meet the criterion of being of 'sufficient' quality to be presented to a jury in the absence of any further identification evidence (Attorney General's Reference, No. 2 of 2002, 2003). As discussed in Chapter 3.5, the video images would also easily attain the recommended Home Office standard in the UK for accurate unfamiliar person identification (Aldridge, 1989; 1994). Some of these videos were taken only a few minutes prior to the identification sessions. Bruce et al. (1999) acquired the close-up photographs and videos for their studies on the same day and the *target present - immediate* condition in the present experiment can be considered analogous to their 'Gold Standard' in a live actor context. Performance in this condition was compared with equivalent footage taken a week prior to the sessions (*target present – time lapse*). The actors were asked to shave facial hair, and they brushed their hair slightly (without cutting it) to simulate minor differences that may occur from day-to-day in everyday life.

According to the Attorney General's reference (Attorney General's Reference, No. 2 of 2002, 2003) if a jury is invited to compare the defendant with a perpetrator in a video, a contemporary photograph of the accused should also be available. Therefore, a further feature was to compare performance when the actors were physically present with when they were shown in close-up photographs. Full-face photographs of the two target actors were taken at the same time as the live identification sessions for use in the equivalent photograph matching trials.

Four novel actors were recruited for this experiment, two acted as targets and two as matched distracters. It would obviously be relatively easy to recruit identical twins, or other close family members for experiments of this type. However, those recruited for this experiment were unrelated, although they were selected by acquaintances of the experimenter for their similarity of appearance. Their photographs were entered into the database described in Chapter 3.8 and rated for similarity by pilot participants to provide a quantitative measure of their resemblance to one another.

The specific predictions for this experiment were that higher errors would be associated with the *target present – time lapse* condition in contrast to the *target present - immediate* condition as the appearance of the actors had been deliberately altered between sessions. In addition, consistent with previous studies, errors were also expected in the *target absent – distracter* condition. However, there were no specific predictions as to whether there would be a difference in the proportion of target absent and target present errors, as this was expected to depend on the facial similarity of the actors involved.

Finally, studies comparing eyewitness memory with lineups made up of live actors or of photographs have found conflicting results. In some, recognition was more accurate with photographs (Dent & Gray, 1975, cited in Dent 1977); in others, an advantage for live actors was found (e.g., E Brown et al., 1977; Cutler & Fisher, 1990; Egan et al., 1977). However, in the remainder no differences were found (Cutler et al., 1989; Shepherd et al, 1982). Due to this inconsistency in recognition studies, no specific predictions were made concerning differences in performance between photographic and live presentation mode in this matching study.

### 7. 1. 2. Method

#### 7. 1. 2. 1. Participants

Participants were 99 male and 277 female adult students, staff and visitors (Mean age = 26.82, *SD* = 10.49) to Goldsmiths College, University of London. None of the participants had taken part in any of the previous experiments or pilot studies.

### 7. 1. 2. 2. Materials and actors

Four white Caucasian male actors aged 20 – 21 years, height 1.83m – 1.87m, weight; 60kg – 73kg, not involved in any previous experiments were recruited for this study (Chapter 3.2). Two of the actors attended target present identification sessions and were videoed twice, once on the day of the identification sessions for the target present *immediate* condition, once approximately a week earlier for the target present *time lapse* condition. The other two recruits acted as distracters in target absent sessions and were videoed once only. As described in Chapter 3.3.1, all were filmed with facial images taking up approximately three-quarters of the screen while turning slowly from left-to-right profile and back for approximately 20-sec. For presentation purposes, playback was in a continuous loop. Still images from these videos are presented in Appendix B.

The photographs of the two targets were black-and-white full-face views, taken at approximately the same time as the identification sessions. Photographs were also obtained of the two distracters and all four are displayed in Figure 7.1. For this experiment, Actor 43 was matched with Distracter 44, Distracter 45 with Actor 46.

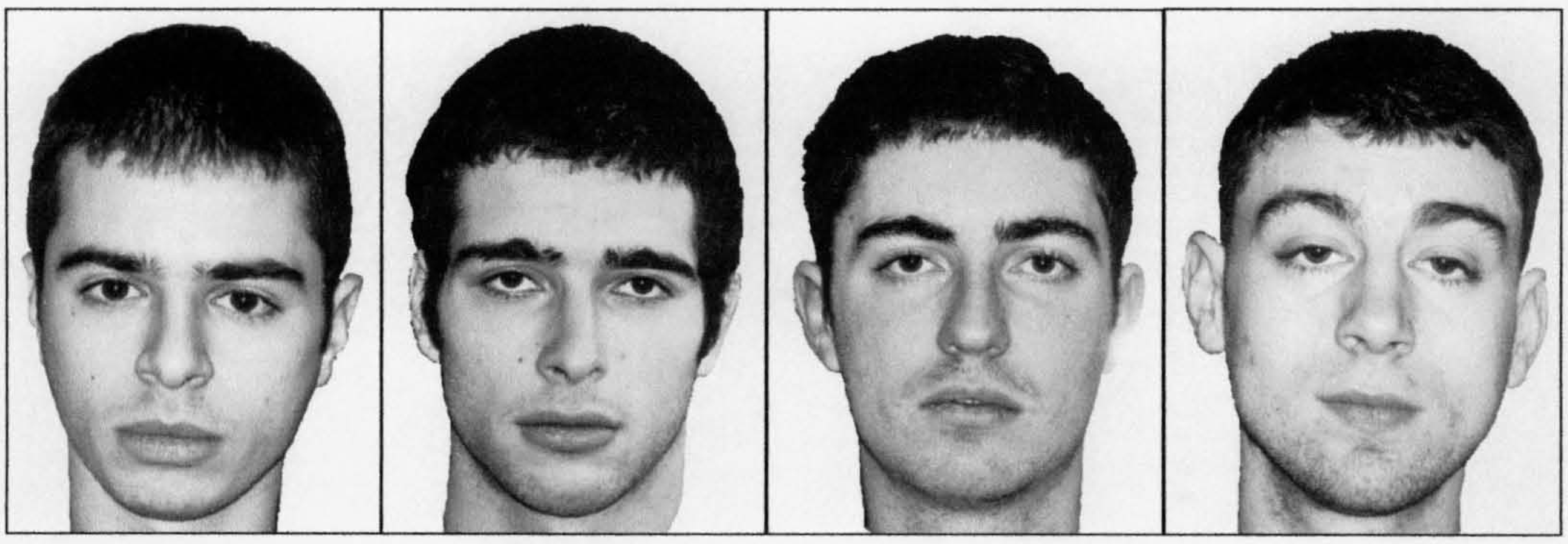


*Figure 7.1: Photographs of the 4 actors recruited for Experiment 7.1 (from left to right: Actor 43; Distracter 44; Distracter 45; Actor 46)*

Unlike the previous experiments reported in this thesis, the actors in this study were not recruited from within a specific social group. However, their photographs were entered into the database of 100 images described in Chapter 3.8 and rated for similarity in a matrix by the 75 pilot participants. From this procedure, two specific faces were paired 25 times, the highest frequency in the matrix. Of the actors involved in this experiment, Actor 43 was paired with Distracter 44 13 times out of 75; Distracter 45 was paired with Actor 46 11 times. Twelve of the 4950 cells in the matrix contained frequencies of 14 or more illustrating that the actors in Experiment 7.1 had a strong physical resemblance but were not uncommonly alike.

#### 7. 1. 2. 3. Design

This experiment utilised a 2 (presentation mode) x 3 (video condition) x 2 (live actor) independent measures single-item identity-verification design. Participants viewed video footage and made a matching decision under one of two conditions. In the 'live' mode they had to decide whether the actor present in person was depicted in the video. In the 'photograph' mode, participants made the same decision to photographs. The image was either of the same actor taken a few minutes prior to the identification session (target present – immediate); the same actor taken approximately a week earlier (target present – time lapse), or was of a matched distracter (target absent - distracter).

Two different live actors were recruited and paired on the basis of facial similarity with two separate distracters. Each of these target actors was treated as a separate level of a third factor. The dependent variables were unadjusted scores on the 8-point identity scale as well as accuracy and confidence measures obtained from score conversion.

#### 7. 1. 2. 4. Procedure

The 'live actor' identification sessions were conducted with participants tested individually or in small computer workshop lectures of between 15 and 20. In the group sessions, the video footage was displayed on an individual computer monitor for each participant. The actor would enter and stand at the front of the room prior to presentation of the footage. They would stand with arms folded and were asked to initially look ahead and keep a neutral expression. The experimenter would ensure that there was no corroboration between participants. In individual sessions, the actor would walk into the room and stand in the same manner, but this time the video was displayed on an individual laptop computer. No specific record was kept as to which equipment individual participants were allocated. Therefore, it was not possible to exactly examine whether responses were biased in any way from differences in this procedure. However, *approximately* 10 out of the 30 participants in each live actor experimental condition were allocated to use the laptop, so it is unlikely that the overall conclusions were greatly biased by this procedure as numbers were fairly consistent.

In the photograph mode, participants were handed one of the two target actor images shown in Figure 7.1, sized A4 for comparison while viewing the footage on the laptop computer.

Participants viewed the footage of one of the target actors or of their matched distracter and responded using the 8-point identity-decision scale used in the previous experiments. Participants were encouraged to ask the live actors to turn their faces to aid their identification decisions. There was no restriction on the number of times each participant could view the footage before making a decision. Full performance feedback was provided at the end of the experiment.

### 7. 1. 3. Results

Table 7.1 displays the mean unadjusted scale data and percentage error rates as a function of video condition, presentation mode and target actor.

#### 7. 1. 3. 1. Unadjusted identity-decision scale data

A 2 (presentation mode) x 2 (actor) x 3 (video condition) ANOVA on the unadjusted scale scores found that the main effect of presentation mode was not significant, $F(1, 364) = 2.60$, $p > .1$. However, the main effect of actor was significant, $F(2, 364) = 18.16$, $p < .001$, scores were higher to Actor 46 than to Actor 43 ($M = 5.71$, $SD = 2.36$ vs. $M = 4.80$, $SD = 2.46$ respectively), indicating a bias towards responding that Actor 45 was present in the video. The effect of video condition was also significant, $F(2, 364) = 63.91$; $p < .001$. Post-hoc Bonferonni tests revealed significantly higher scores in the immediate condition than in the time lapse condition, which were also significantly higher than in the distracter condition ($M = 6.60$, $SD = 1.93$ vs. $M = 5.48$, $SD = 2.20$ vs. $M = 3.67$, $SD = 2.27$ respectively; $p < .001$ for all comparisons).

*Table 7.1: Mean unadjusted scale scores and percentage error rate to each actor as a function of video condition and presentation mode (standard deviations in parentheses)*

| | | Target Present Immediate | | Target Present Time Lapse | | Target Absent Distracter | |
|---|---|---|---|---|---|---|---|
| Presentation Mode | | Actor A43 | Actor A46 | Actor A43 | Actor A46 | Actor A43 | Actor A46 |
| Live Actor Mode | *n* | *32* | *31* | *32* | *30* | *32* | *31* |
| | Mean Scale Score | 5.94 (2.18) | 6.97 (1.85) | 5.63 (2.12) | 6.37 (1.75) | 3.25 (2.24) | 4.42 (2.31) |
| | Percentage error rate | 18.7 | 12.9 | 34.4 | 16.7 | 31.2 | 51.6 |
| Photograph Mode | *n* | *32* | *31* | *31* | *31* | *32* | *31* |
| | Mean Scale Score | 6.59 (1.86) | 6.94 (1.67) | 4.23 (2.30) | 5.71 (2.10) | 3.16 (1.83) | 3.87 (2.51) |
| | Percentage error rate | 21.9 | 12.9 | 51.6 | 29.0 | 25.0 | 29.0 |

The interaction between presentation mode and video condition was also significant, $F(2, 364) = 3.25$, $p < .05$ (Figure 7.1.1). Bonferonni simple effects analyses revealed that in the photograph mode, $F(2, 185) = 37.70$, $p < .001$, the pattern of results was consistent with those reported for the main effect of video condition above ($p < .01$ for all Tukey's post-hoc tests). The simple effects of video condition were also significant within the live presentation mode, $F(2, 185) = 27.04$, $p < .001$. However, Tukey's post-hoc tests found that scores did not significantly differ in the target present time lapse and immediate conditions ($p >$ .1), but both sets of scores were higher than in the distracter condition ($p < .01$ for both comparisons). The remaining two-way interactions and the three-way interaction were not significant ($p > .1$).

### 7. 1. 3. 2. Accuracy data

The percentage of errors in each condition is presented in Table 7.2. Overall, 27.9% of responses were incorrect; 24.8% in target present conditions, 34.1% in target absent conditions. A 2 (presentation mode) x 2 (actor) x 3 (video condition) ANOVA conducted on the error rate data found that the main effects of presentation mode, $F < 1$ and of actor, $F(1, 364) = 1.28$, $p > .1$ were not significant. However, the effect of video condition was significant, $F(2, 364) = 6.31$, $p < .005$. Post hoc Bonferonni tests revealed significantly fewer errors in the immediate condition (16.7%, $SD = 37.4$) than in the time lapse (33.1%, $SD = 47.2$) and distracter conditions (34.1%, $SD = 46.6$, $p < .01$ for both comparisons). There were no differences between the time lapse and distracter conditions ($p > .1$).

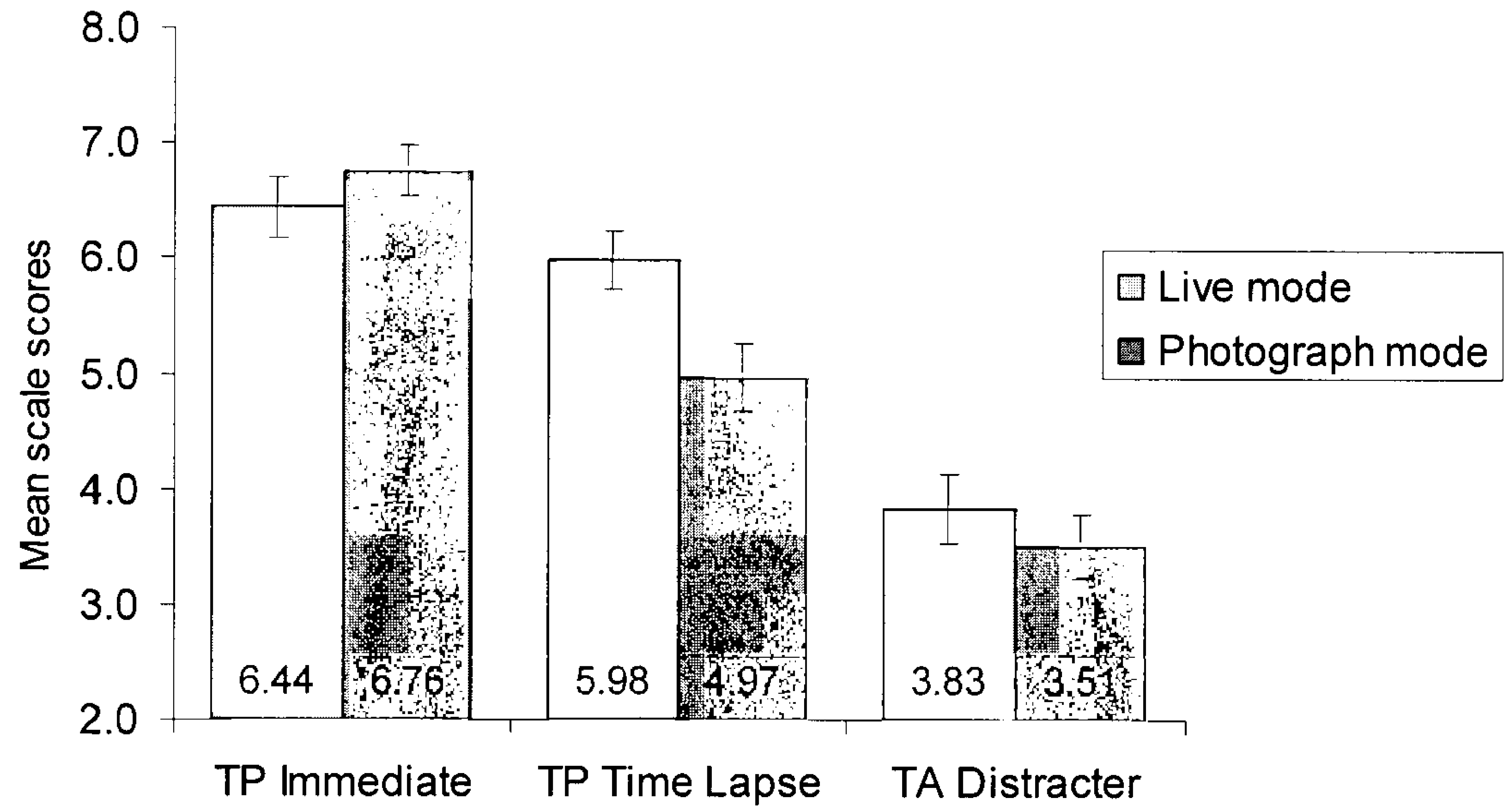



*Figure 7.2: Mean scale scores as a function of presentation mode and video condition (error bars denote standard error of the mean)*

The interaction between presentation mode and condition was also significant, $F(2, 364) = 3.49$, $p < .05$ and is displayed in Figure 7.1.2. Bonferonni-corrected simple effects analyses revealed that there was a difference in accuracy across video conditions when actors were live, $F(2, 185) = 5.36$, $p < .01$; Tukey's paired comparisons found that although there was no significant difference in error rates in the immediate and time lapse conditions ($p > .1$), significantly more were made in the distracter condition ($p < .05$). The simple effects of video condition in the photograph mode was also significant, $F(2, 185) = 4.18$, $p < .05$. Tukey's tests revealed that error rates were similar for the time lapse and distracter conditions ($p > .1$) with marginally fewer made in the immediate condition ($p < .1$).

There was also a significant interaction between actor and condition, $F(2, 364) = 4.33$, $p < .05$; Bonferonni-corrected t-tests revealed that error rates were similar for both actors in the immediate and distracter conditions ($p > .1$ for both analyses). However, more errors were made to Actor 46 than to Actor 43 (42.8 vs. 22.9% respectively) in the time lapse condition, $t(122) = 2.39$, $p < .05$ suggesting that Actor 46's appearance substantially changed in the intervening period. The three-way interaction was not significant, $F < 1$.

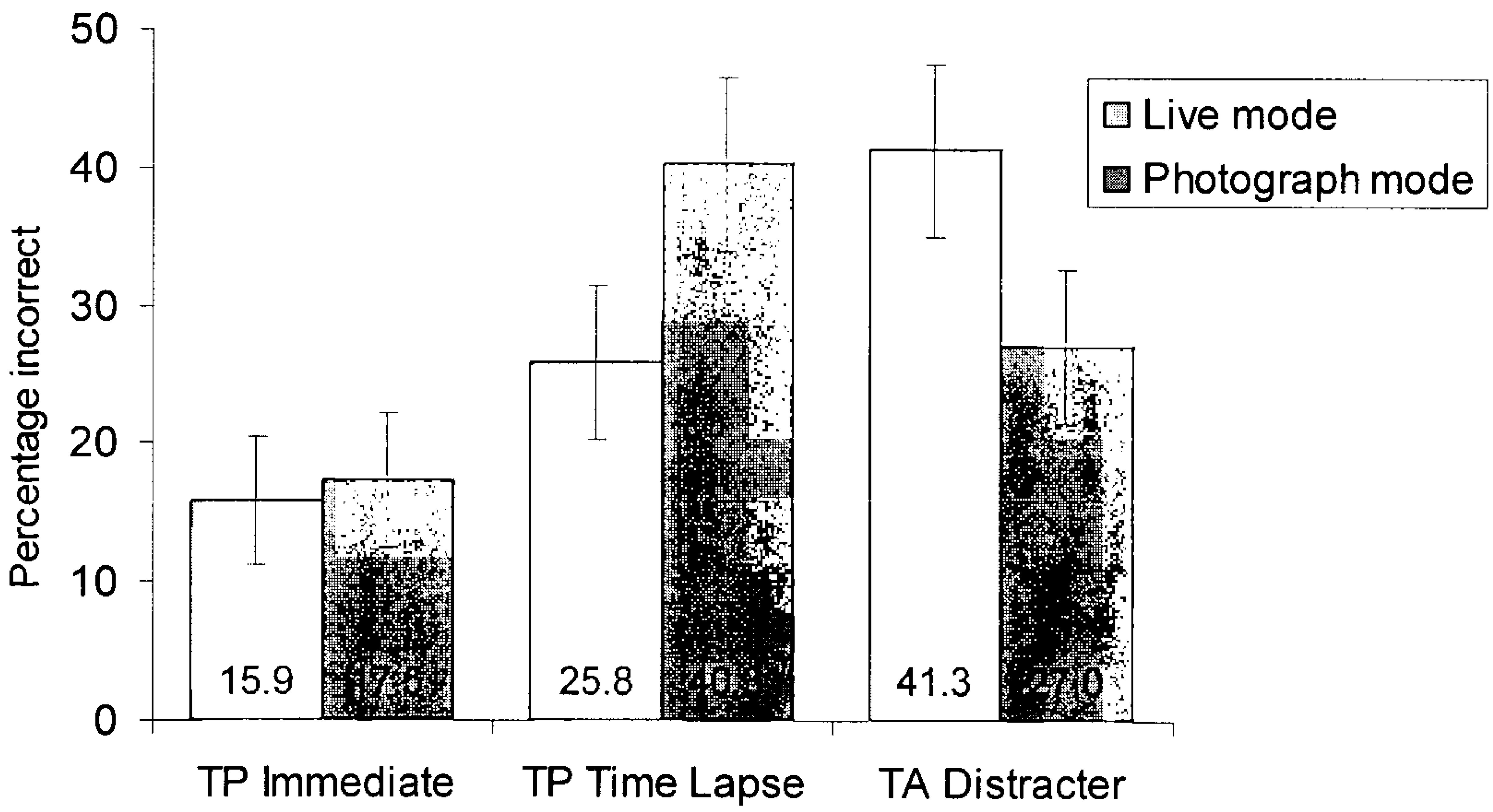



*Figure 7.3: Mean percentage error rates as a function of presentation mode and video condition (error bars denote standard error of the mean)*

### 7. 1. 3. 3. Confidence levels

Scale scores were converted to produce confidence data. However, the data from the trials involving the two different actors were pooled as cell numbers were low in some conditions. A 2 (accuracy) x 2 (presentation mode) x 2 (decision type) ANOVA conducted on these data revealed that the main effect of accuracy was highly significant, $F(1, 368) = 24.25$, $p < .001$; correct decisions were associated with higher confidence than incorrect decisions ($M = 3.02$, $SD = 1.06$ vs. $M = 2.32$, $SD = 0.94$ respectively). The effect of presentation mode was not significant ($p > .1$). However, the effect of decision type was significant, $F(1, 368) = 5.63$, $p < .01$; 'same' decisions were associated with higher confidence than 'different' decisions ($M = 3.00$, $SD = 1.03$ vs. $M = 2.54$, $SD = 1.09$ respectively). None of the interactions were significant ($p > .1$).

### 7. 1. 4. Discussion

Experiment 7.1 verified that even in the most optimal conditions, unfamiliar face matching from video images to either photographs or to actors physically present is error prone. Furthermore, participants were inappropriately confident when making these decisions. When the footage was a week old in the target present – time lapse condition, 25.8% of participants wrongly believed that the actors shown in video were not present in person. Even more were incorrect when the same actors were depicted in photographs (40.3%), with over 50% mistaken in trials involving Actor 43. This effect may be explained to some extent in that both actors had been asked to alter their hairstyle and to shave their facial hair. The still full-face photograph probably highlighted these changes as decisions could be made on the basis of a single viewpoint only. When targets were present in person the extra visible cues appear to have assisted decision making, reducing the number of errors.

Approximately 17% of participants also made errors in the target present - immediate condition when the footage had been taken less than an hour previously and the actors had been instructed not to change their appearance apart from wearing different clothing. Indeed, the error rates in this condition were approximately the same regardless of target actor or presentation mode. These false negative results illustrate that even in 'Gold Standard' conditions, with minimal task demands, unfamiliar face matching is unreliable. Indeed, the consequences could be serious if replicated in forensic investigations, as a guilty suspect would evade justice. However, to some extent these errors in the target present immediate condition may reflect a natural cautiousness, from the mainly psychology undergraduate participants when asked to carry out what is perceived to be a straightforward task. Indeed, psychology students routinely have their perceptions tested and occasionally 'tricked'.

However, if the participants in this experiment were 'suspicious' of experimenter motives, this would naturally result in a tendency to respond using a conservative criterion that actors were not depicted in the videos. However, if a conservative response bias of this type was operating, which should have an overall positive effect on identification accuracy when targets are absent, the results of the target absent – distracter trials are of even greater concern. When the data from both actors is combined, over 40% of participants wrongly believed live actors were depicted in video when in fact a distracter was shown. Indeed, over 50% of participants mistook Distracter 45 for Actor 46. Performance was better if the actors were depicted in photographs rather than being physically present. However, more than one-in-four of the participants still made incorrect decisions. The actors in Experiment 7.1 had been specifically recruited for their similarity of appearance. Nevertheless, the results of the similarity matrix pilot study illustrate that even though these actors were rated as highly similar, other pairs of faces within this set were even more alike.

Consistent with previous experiments in this thesis, accuracy in decisions was associated with a higher level of confidence. However, in contrast to the results of all previous experiments using the medium-range footage, confidence in 'same' decisions was higher than for 'different' decisions. Taken with the number of incorrect responses in the target absent conditions, these results suggest that with high resolution images, participants were more assured that their perceptions concerning a matched identity could not be mistaken. However, when images are of lower quality as in the previous experiments, confidence in positive identifications was much lower. From a legal perspective, if replicated by jurors in court basing their verdicts on identification in high-quality video footage, these findings suggest that an innocent defendant, strongly resembling the real offender would have little chance of avoiding a conviction. However, it also suggests that if jurors have doubts about the identity of the perpetrator shown in video, a guilty defendant may escape punishment unless quality is high.

Even though no specific limits were set, in all of the experiments reported so far in this thesis, participants have had a relatively limited viewing time in which to make their decisions as to the identity of the actors. If they were acting as jurors in a trial, it is likely that the longer exposure they would have of the accused would increase their familiarity with that individual. Indeed, in Australia, jurors are particularly encouraged to compare the defendant with video evidence, and most identification testimony from police officers is prohibited, as jurors are considered likely by the time of deliberation to have had more exposure to the defendant (*Smith v The Queen*, 2001).

Even when using low-quality images, recognition of familiar people in this type of task has been found to be very accurate. This would suggest that jurors would be less likely to make errors of identification, if a trial proceeds for more than a few hours. However, although some research has been directed at the question of how faces are learned (Bonner et al., 2003a; Bonner, Burton, Jenkins & McNeil, 2003b; Bruce et al., 2001; Clutterbuck & Johnston, 2002; 2005; Ellis et al., 1979; O'Donnell & Bruce, 2001), it is unclear how much exposure is required to ensure maximum performance.

There is evidence that the learning of faces may be enhanced through semantic knowledge about the person (Bonner et al., 2003b), or if encouraged to focus on the personality of an individual in a social setting (Bruce et al., 2001; Experiment 3). Indeed, an eyewitness is given more credence in court if they have had regular past social contact with a defendant (e.g., *R v Grimer*, 1982). However, identification testimony may also be admissible if a witness claims to have familiarised him or herself to an offender, purely from viewing video evidence (e.g., *R v Clare and Peach*, 1995). This scenario is examined in the following chapter.

# Chapter 8: The familiarisation of facial images in video

## 8. 0. Introduction

Empirical studies have regularly found that in contrast to unfamiliar faces, the recognition of familiar people tends to be highly accurate, even if image quality is extremely poor (e.g., Bruce, 1982; Bruce at al., 2001; A.M. Burton et al, 2001). Models of face recognition generally specify that the processing of familiar and unfamiliar people involves different operations (e.g., Bruce & Young, 1986; A.M. Burton et al., 1999) and neuropsychological research has implicated separate functional areas within the brain (e.g., Andreasen, O'Leary, Arndt et al., 1996; Malone, Morris, Kay & Levin, 1982; Young, Newcombe, de Haan, Small & Hay, 1993). The manner in which faces are perceptually processed is also determined by their familiarity. External facial features such as the chin and hairstyle are equally, if not more important than internal features in the recognition or matching of unfamiliar or newly-encountered people (Bonner et al., 2003a; Bruce et al., 1999; Clutterbuck & Johnston, 2002; 2005; Newcombe & Lie, 1995; Young et al., 1985). Conversely, internal features (e.g., the eyes, mouth and nose) are more critical when faces are familiar (Clutterbuck & Johnston, 2002; 2005; Ellis et al., 1979; O'Donnell & Bruce, 2001; Young et al., 1985).

These factors impact on evidence in court. There may have been no witnesses at an incident, but CCTV images may have been captured. The testimony of an individual making a positive identification of a familiar offender from footage may be treated similarly to that of an eyewitness actually present (e.g., *R v Caldwell and Dixon*, 1993; Attorney General's Reference, No. 2 of 2002, 2003). A witness recognising someone highly familiar will also have more credence than if the image is of someone relatively unfamiliar (e.g., *R v Grimer*, 1982). The ability to extensively replay video evidence may give even greater weight to such testimony.

However, the evidence of an individual previously unfamiliar with a suspect can also be admissible if they have conducted extensive viewing of CCTV images in order to familiarise themselves with any offenders (e.g., *R v Clare and Peach*, 1995; Attorney General's Reference, No. 2 of 2002, 2003). In one case (*R v Clare and Peach*, 1995), a police officer viewed CCTV evidence frame-by-frame at least 40 times and identified the defendants when cross-referencing to separate photographs. This involved the extraction of a number of stills, slow motion analyses and comprehensive inspection of the images. From this, the prosecution claimed the officer was able to provide detailed identification testimony, which, without this extensive examination would not have been possible. His testimony was cross-examined, and his status was considered that of an 'ad-hoc' expert witness.

A similar report was given in a case referred to the Attorney General (Attorney General's Reference, No. 2 of 2002, 2003). A member of a police video viewing team, previously unfamiliar with the defendant spent a "considerable number of hours viewing the film, and, in consequence, became familiar with the appearance of persons to be seen in it". He later recognised the defendant by a 'chance' encounter, and appeared as a witness on this basis.

However, it is unclear under what circumstances exposure of a previously unfamiliar face is sufficient for that face to be categorised as familiar and for such testimony to be reliable. Indeed, in Canada, the Court of Appeal ruled that similar evidence from police officers would not be admissible due to a lack of previous knowledge of the defendant and the potential for a lack of impartiality (Leaney & Rawlinson, 1988, cited in Mead, 1998). It is also unclear whether it is possible to sufficiently 'learn' faces in such an impoverished manner, in the absence of any form of social interaction or knowledge of personality characteristics.

Bruce and Young (1986) in their structural model of face recognition, suggest that repeated exposure to a novel face across a variety of poses, expressions or distances is necessary for an internal representation, or face-recognition unit (FRU) of that face to be developed. This is strengthened by repeated exposure. When that familiar face is later perceived, a number of FRU's may respond. However, only the FRU associated with the correct target normally achieves a threshold level exceeding all others. Errors in recognition may occur, as the threshold of an inappropriate FRU may be surpassed, if, for instance, the target face is seen fleetingly from an unusual angle. However, this is less likely with a highly familiar person. According to this model, initially weak FRU's are formed on exposure to unfamiliar faces. However, unfamiliar face recognition involves extracting viewpoint-specific information based on pictorial elements rather than on a viewpoint-free representation. Thus, unfamiliar face recognition can easily be disrupted by stimuli changes and a superficial similarity may result in a mistaken identification.

Some indication of how novel faces become familiarised is provided by past research (e.g., Bonner et al., 2003a; Bruce et al., 2001; Clutterbuck & Johnston, 2004; 2005; O'Donnell & Bruce, 2001), with relatively brief exposure to faces modifying performance on some tasks. For instance, Clutterbuck and Johnston (2004) demonstrated that viewing full-face images over ten presentations, each lasting two-sec, facilitated response times to alternative (three-quarter) views of the same faces in a later gender-decision task, suggesting the development of rudimentary viewpoint-free FRUs. Furthermore, Bonner et al. (2003b) found that daily exposure to 24 high-quality close-up faces in 90-sec video clips for three days altered the manner in which the same faces were differentially recognised using external or internal features. The authors argue that this demonstrates the preliminary establishment of permanent face representations and suggest that further training would have resulted in the typical familiar face internal feature advantage.

Although these findings have theoretical interest, viewing of CCTV images will normally involve scrutinising full faces, unless features are obscured or disguised. An experiment by Bruce et al. (2001; Experiment 3) examined face learning in this context. Participants were presented with a series of 30-sec moving high-quality video clips showing faces in close-up. Some watched the footage in pairs and were asked to "chat about the faces between yourselves as you watch the video" (p. 215).

Matching to the same targets in target present arrays was more accurate (98%) than controls (88%) who had viewed the videos in isolation. The authors suggest that participants in the paired condition discussed perceived personality elements in the faces during the socialisation procedure, inducing a deeper level of processing. However, when the targets were absent from the arrays, the social-familiarisation group made errors in 32% of trials, each involving the selection of an incorrect face. Although accuracy in this condition was superior to the isolation group (49%), these error rates would still be unacceptable from a forensic perspective.

The relative proportion of target present and target absent errors in the Bruce et al. (1999) experiment indicate that participants had a relatively liberal matching criterion, inducing a high level of false positive responses. Furthermore, there may be an initial advantage in learning faces from social interaction, but intensive viewing of images for longer in isolation may eliminate this benefit. Nevertheless, there are no indications of equivalent procedures from the legal reports of cases in which eyewitnesses have attempted to familiarise themselves with video images (e.g., *R v Clare and Peach*, 1995; Attorney General's Reference, No. 2 of 2002, 2003).

**Experiment: 8. 1. Familiarisation to faces shown in video**

Experiment 8.1 was instigated to examine whether it is possible to be sufficiently familiarised to individuals from extensive viewing of video images alone, in order to provide reliable identification testimony. More specifically, the aim was to investigate whether participants in a *learning* condition, required to replicate some of the publicised procedures described in *R v Clare and Peach* (1995), would be better at accurate identifications than those in a *control* group who viewed the same footage for a limited period, as would be expected of a jury. This would test the assumptions that such a witness could develop an 'ad-hoc expertise' in recognising those particular faces.

Half of the participants were allocated to the *learning* condition, viewing the medium-range video images of 12 different actors, described in Section 3.1, over three one-hour sessions conducted on separate days. Performance was compared with the *control* group who viewed the same footage in one session only. During the course of the experiment, both groups made a series of matching decisions to arrays of photographs to ensure they were attending to the footage. As such, the procedure followed by the learning group could be conducted by police officers matching images to a series of mug-shot photographs. Furthermore, throughout the course of the experiment, arrays and videos depicted faces from alternative viewpoints or in disguise to encourage the development of viewpoint-free facial representations, as theorized by Bruce and Young (1986).

In addition, half of the arrays for the first sessions for both groups were target absent, to replicate a situation that a police officer may encounter, if, for example, the actual perpetrator is not initially a suspect. However, in the final trials, the target actors were added to arrays, initially without informing the participants. The final assignment was a two-alternative forced choice task in which a photograph of the actor shown in the video was always present alongside a matched distracter. This distracter was the actor incorrectly identified most often by each individual participant throughout their previous trials. This could be experienced by a police officer who had consistently identified an innocent suspect as being on video and was then finally confronted with a photograph of the real offender.

The predictions of this experiment were that participants in the learning condition would be more accurate than the control group in the final trial, having familiarised themselves to the actors shown on video. A secondary aim was to investigate the potential action of an incorrect response bias in target absent conditions. In some cases, participants were expected to make consistent, incorrect selections of the same distracter face across a series of trials showing that face from different viewpoints, possibly indicating from the theoretical perspective of Bruce and Young (1986), the establishment of an inappropriate FRU. Therefore, for all participants, it was expected that target present performance would be better than target absent performance. However, due to being exposed to proportionally fewer target absent trials, this effect was expected to be less robust in the control group.

### 8. 1. 1. Method

#### 8. 1. 1. 1. Participants

Participants were first year adult undergraduate students (7 male; 37 female) at Goldsmiths College, University of London (Mean age = 21.98, *SD* = 5.63). All gained course credit for participation. Assignment to experimental group was self-selecting as participants either signed up to take part in three 1-hour sessions as a member of the learning group, or for a single session for the control group. None of the participants had taken part in any of the previous experiments or pilot studies.

#### 8. 1. 1. 2. Design

This study employed a 2 (experimental condition) x 2 (target presence) mixed face identity-verification design. Participants repeatedly viewed a series of videos depicting 12 actors. The *experimental condition* independent measures variable had two levels. As such, the *learning* group, participated in three 1-hour sessions, designed to replicate procedures conducted by ad-hoc expert witnesses. In contrast, the control group completed a single session only, designed to be more like the experience of a juror.

For the learning group, each session involved participants continually viewing the series of videos depicting the target actors and on each consecutive viewing they were required to match the target to their image within arrays of photographs. Array sizes, the position of the targets and distracters within arrays, viewpoint of the photographs and whether the target on video was in disguise or not varied on each trial. In contrast, the *control* group viewed the images a minimal number of times only, ensuring that they comprehensively understood the task demands, without developing more than a basic familiarisation to the targets.

The repeated measures variable, *target presence* also had two levels. Videoed actors were present or absent in the photograph arrays during the initial series of trials experienced by both groups. For this, for half the participants, the photographs of six of the 12 actors were present in every array. The remaining six were always absent. This variable was fully counterbalanced so that actors were shown an equal number of times in target present and target absent conditions.

For the final trial, conditions were similar for both the learning and the control groups. As such, the targets in video were always present in one of two photographs and the primary dependent variable was matching accuracy using a two-alternative forced-choice identification-matching task.

**8. 1. 1. 3. Materials**

The 36 videos used were of 12 of the 24 actors, depicted in the three different disguise conditions described in Section 3.3. Fifteen different randomly arranged photographic arrays were constructed to display alongside each of these 12 actors. This involved six different array types, each depicting faces from alternative viewpoints. Two array booklets (paper sized A4) were produced for each experimental session so that targets were presented equally often in target present and target absent conditions. Full details of arrays are described in Section 8.1.1.4.

In target present conditions, the actor shown in the video was displayed within the arrays. In target absent trials, this photograph was replaced by that of a further distracter. Distracter photographs for all arrays were from the database of 100 facial images described in Section 3.8 and were based on the selections by pilot participants as being most similar in appearance to the target. Due to the limited number of photographs in this database some distracters appeared in the arrays of more than one target actor. However, photographs of the 12 target actors were never used as distracters[1]. Response sheets were provided, so that participants could circle a letter of choice responding to the faces in each array and unless otherwise stated, included a 'not present' selection option.

### 8. 1. 1. 4. Procedure

#### 8. 1. 1. 4. 1. Overview

This section gives an overview of the conditions experienced by participants. However, full details are provided below. Half of the participants signed up to participate in the three-session *learning* condition, designed to replicate features of the evidence provided by ad-hoc expert witnesses in court. For this, a series of matching tasks from video to photograph arrays were conducted in order to induce increasing familiarisation with the 12 videoed actors. Consecutive photographic arrays depicted the faces from alternative viewpoints to encourage the development of viewpoint-free facial representations. In addition, the medium-range video clips depicting the actors in the three disguise conditions were employed (hat, dark glasses and no disguise; see Section 3.3), designed to encourage learning using different facial features, dependent on their availability in the images, as, for instance, the eyes would be obscured when the actors were depicted in glasses and the hair obscured by the hat.

To help them make decisions, full training was provided to the learning group on how to use the video equipment controls, such as rewinding, viewing in slow motion and how to pause the tape for the extraction of stills, features of the evidence described in *R v Clare and Peach* (1995).

The remaining participants were members of the *control* group, designed to replicate conditions experienced by a jury and they therefore encountered some of the above conditions in a single session only.

---

[1] Examination of all results found that no participant selected the same distracter in more than one trial in any phase of the experiment, indicating that the use of repeated distracters did not appear to impact on the results.

All participants were informed that they should attempt to familiarise themselves with the videoed actors and that in some arrays the target would not be present, although no indication was given of the likelihood of occurrence. In fact, half of the arrays for the first two 1-hour sessions for the learning group, and for the first of three stages for the control group were target absent.

In the concluding trials of both experimental groups, the target actors were added to arrays, initially without informing the participants. The final assignment was a participant-specific two-alternative forced choice task in which a photograph of the actor shown in the video was always present alongside a matched distracter.

Participants were provided with a booklet at the start of each session containing the facial arrays. The 12 actors on video were always shown in the same order.

**8. 1. 1. 4. 2. Session 1 (Learning Group)**

The first session involved three separate stages during which participants viewed the 12 videoed actors a minimum of ten times each. Viewing time was approximately one hour. Videos were played on a 26" colour television with participants sitting approximately 1m from the screen.

In **Stage 1,** participants were presented with a video tape depicting two consecutive clips of the 12 actors in no disguise. A page in the array booklet showing six left three-quarter facial photographs was specific to each actor. Using the response form, participants were required to select from the array the actor they believed was present in the respective video, or to respond 'not present'.

In **Stage 2**, participants were provided with a second video. This showed each actor six times in succession, twice in each of the three disguise conditions in the following order: no disguise, glasses, hat, hat, glasses, and no disguise. For each actor on this second video, participants had to choose faces from two different arrays. The first depicted six right three-quarter faces. The second depicted ten full-face images. Participants were trained on and encouraged to utilise the controls of the video player such as slow motion, freeze frame and rewind.

**Stage 3**, was a repeat of Stage 1, only each array was made up of six right three-quarter facial images.

#### 8. 1. 1. 4. 3. Session 2 (Learning group)

The second session took place a minimum of 24 and a maximum of 96 hours later, with each actor again being viewed on video a minimum of ten times.

**Stage 4** replicated Stage 1, except each array was randomly displayed in a different arrangement.

In **Stage 5** participants were provided with the same video as in Stage 2 and a response booklet containing 72 pages, each depicting six different arrays for each of the 12 actors. The order of arrays was as follows: - right profile (six faces); left three-quarters (six faces); full-face (ten faces); left profile (six faces); right three-quarters (six faces) and full-face (ten faces).

**Stage 6** was a replication of Stage 3, except each array was randomly displayed in a different arrangement.

#### 8. 1. 1. 4. 4. Session 3 (Learning group)

The third session took place a minimum of 24 hours after the second and a maximum of one week following the first.

In **Stage 7**, participants were required to extract a minimum of two digital stills from each clip of the actors shown in no disguise (resolution 352 x 240 pixels) using Dazzle Digital Video Creator software, presented on a 16" computer monitor. They were instructed that at least one still should depict that particular actor from a frontal view and to use the various facilities of the video player to extract additional stills if required. Participants were then presented with two arrays for each actor. The first array depicted six right three-quarter views, the second array: - ten full-face images. For the first time in the study, *all* arrays were target present. However, participants were not informed of this and the 'not present' response option was still available. They were encouraged to utilise any of the stills collected of a particular actor, and if they wished to enlarge the obtained images on the screen.

The same video used in Stage 7 was provided in **Stage 8**. Arrays were target present, with each array displaying two photographs of each of the six faces. One photograph was a left profile view; the second was a full-face view. In this stage, the 'not present' option was unavailable.

In **Stage 9** a set of 12 pairs of images for the final two-alternative choice trials showing faces from a right three-quarter view was specifically constructed for each individual participant, dependent on previous matching decisions by that participant. One of the pair always depicted the target actor. The second depicted the distracter selected by that participant most often in previous trials. If two or more distracters had been equally selected, the face was of the distracter chosen most recently. If no incorrect decisions had been made, the second image was that rated by pilot participants as most similar in appearance to the target.

**8. 1. 1. 4. 5. (Control group)**

The control participants took part in one session of approximately one hour only, viewing the videos a maximum of six times each, and completing exactly the same matching trials to those described for the learning group at Stages 1, 8 and 9. At the commencement of the session they were also informed that they would be familiarizing themselves with the actors shown in the videos.

Full performance feedback was provided to all participants at the end of their final session.

## 8. 2. Results

A large quantity of data was collected for this experiment. However, of interest was the comparison of performance between the learning and control groups at critical phases in which the same conditions were experienced. The presentation of the arrays had been fully counterbalanced to ensure that each target actor was shown an equal number of times in exclusively target present and target absent trials in the preliminary stages. Initially entered as a further factor in all analyses reported below, this counterbalancing variable did not interact with any other and therefore data was pooled when reported below ($p > .1$).

### 8. 2. 1. Examination of Baseline Performance

The first two sets of analyses examined the baseline trials conducted by both groups at Stage 1, and additionally for the learning group following the first two one-hour sessions at Stage 6. The proportion of correct and incorrect responses (expressed as percentages with standard deviations), defined as each possible outcome for target present trials and correct rejections for target absent trials at these stages is presented in Table 8.1.

*Table 8.1: Percentage of correct and incorrect responses in target present conditions; and percentage of correct rejections in target absent trials as a function of experimental stage and group*

| | Target present | | | | | | Target absent | |
|---|---|---|---|---|---|---|---|---|
| | Hits | *SD* | Incorrect | *SD* | Miss | *SD* | Correct rejections | *SD* |
| Learning group | | | | | | | | |
| Session 1 Stage 1 | 45.5 | *16.4* | 36.4 | *12.2* | 18.2 | *14.5* | 45.5 | *24.2* |
| Session 2 Stage 6 | 58.3 | *25.1* | 22.0 | *20.8* | 19.7 | *21.0* | 40.2 | *22.8* |
| Control group | | | | | | | | |
| Session 1 Stage 1 | 38.6 | *20.8* | 34.1 | *19.6* | 27.3 | *18.2* | 56.1 | *22.7* |

Combining the data from both experimental groups, at Stage 1, 41.1% of target present trials involved the correct selection of the target from arrays; and when targets were absent 50.8% of trials involved correct not present decisions. A series of independent-measures t-tests were conducted to compare the learning and control groups on each of the outcomes listed in Table 8.1. The groups did not significantly differ in the percentage of target present hits, $t(42) = 1.21$, $p > .1$, incorrect responses, $t(42) = 0.46$, $p > .1$ or in the target absent correct rejections, $t(42) = 1.50$, $p > .1$. However, there was a marginally significant trend for the control group to make more target present misses, $t(42) = 1.83$, $p = .074$, an indication that slightly more 'not present' responses were made by this group. A possible interpretation of these results is that those in the learning group, being aware that they would be taking part in three separate sessions, were more motivated to attempt to locate a target from the array in comparison to the control group who participated in one session only.

### 8. 2. 2. Comparison of baseline performance with that at the end of Session 2

The second set of analyses was conducted to examine whether the learning group response outcomes had altered following the first two 1-hour sessions. The trial conducted at Stage 6 was a replication of the initial Stage 1 trial, except that the faces in arrays were depicted from a different viewpoint. A series of five repeated-measures t-tests examined whether the overall number of responses of each outcome type made by the learning group from Table 8.1 differed between Stages 1 and 6. In target present conditions, the number of hits significantly increased, $t(21) = 2.40$, $p < .05$, with a subsequent decrease in the number of incorrect distracter selections, $t(21) = 2.72$, $p < .05$, providing some indication of face learning. However, there were no significant changes in the number of target present misses, $t(21) = 0.40$, $p > .1$, target absent correct rejections, $t(21) = 1.09$, $p > .1$ or in the proportion of 'not present' responses, $t(21) = 0.55$, $p > .1$.

### 8. 2. 3. Stage 9 two alternative-choice trials

The final trial was a two-alternative forced-choice task in which one of the pair of right three-quarter photographs always depicted the target present in the video. The second photograph was individually assigned based on each participant's previous preference when confronted with that particular target in video. Using this allocation procedure, in one case almost all of the participants (42 out of 44; 90.9%) received a photograph of the same distracter, whereas in another trial the distracter selected most often was actually viewed by a minority of participants (36.4%). Overall, in almost two-thirds of cases (65.5%), the distracter chosen was the most preferred choice of the majority of participants. Nevertheless, in some cases a specific distracter was presented to a single participant only. Overall, there was a maximum number of five, and a minimum of three distracters presented to different participants when confronted by the same target actor ($M$ = 4.33). However, a 2 (experimental group) x 2 (target presence) mixed ANOVA found no significant effects of distracter number across each experimental condition ($p > .1$).

The proportion of correct responses in Stage 9 (expressed as a percentage) for both groups as a function of whether targets had been present or not in previous phases is presented in Table 8.3. Overall, 72.9% of responses were correct. These data were entered in a 2 (target presence) x 2 (experimental group) mixed ANOVA. The main effect of target presence was significant, $F(1, 42) = 15.16$, $p < .001$; more correct responses were made to actors whose faces had always been present in previous trials ($M$= 79.5%, $SD$ = 17.56) than those whose faces had been absent from arrays in the earlier phases of the experiment ($M$ = 66.3, $SD$ = 14.1). However, both the main effect of experimental group, $F < 1$ and the interaction, $F < 1$ were non-significant.

*Table 8.2: Percentage of correct responses at Stage 9 to targets that had been present or absent in previous phases (standard deviations depicted in parentheses)*

| | Learning group | | Control group | |
|---|---|---|---|---|
| | Target present | Target absent | Target present | Target absent |
| Session 3 | 81.1% | 68.2% | 78.0% | 64.4% |
| Stage 9 | (15.7) | (16.2) | (19.5) | (11.8) |

### 8. 2. 5. Error rates associated with specific targets

Only one participant performed at ceiling in Stage 9 by correctly matching all 12 actors. Furthermore, only one of the 12 actors was correctly identified by all participants (Actor 16). In contrast, approximately only half of the participants correctly identified the target in two other cases (50.0%: Actor 18 and 52.3% Actor 32). Indeed, 36.4% of participants chose one specific distracter (Actor 31) instead of the actual target (Actor 32). Furthermore, five different distracters were selected (Actors 03; 04; 06; 57; 62) in trials involving a further target actor depicted on video (Actor 02), and across all 12 targets a mean of 3.25 different distracters for each were selected in the final stage (see Appendix E for images of actors).

### 8. 1. 3. Discussion

Experiment 8.1 revealed that despite repeated intensive exposure to the 12 actors shown in video, unfamiliar face matching from video to photographs was still error prone. In the final stage, participants made a series of two-alternative forced-choice judgements, as to which of two facial photographs was of an actor simultaneously shown on video. In this phase, approximately one-quarter of selections (27.1%) were incorrect. However, in contrast to expectations, no differences were found between the learning group who viewed each video at least 25 times, taking approximately 13 minutes per clip and a control group who viewed the same 30-sec videos a maximum of six times each.

Moreover, for both groups, performance was worse in conditions in which the target had been absent in initial trials in comparison to when the same target had been constantly present. This result was expected for the learning group, as throughout the experiment, participants were required to make matching judgements while different views of photographs were simultaneously available in arrays. However, only a modest effect of target presence was expected for the control group as exposure to the single target absent stage was comparatively brief. Nevertheless, the effect of target presence in the control group was similar to that of the learning group. There is some evidence that face matching tasks can be moderated by extremely brief repeated exposure to faces (e.g., Clutterbuck & Johnston, 2004). However, it is unlikely that this was due to the development of a strong permanent facial representation and the lack of target presence differences across the experimental groups was unexpected.

From a forensic perspective these results suggest that in contrast to rulings in previous court cases (e.g., *R v Clare and Peach*, 1995; Attorney General's Reference, No. 2 of 2002, 2003) a police officer extensively viewing footage would be no better at making detailed matching judgements to video than an individual juror who may be played the same tape a few times in court. Indeed, only one participant in the learning group accurately identified all 12 targets in the final two-alternative stage. Performance also varied across trials with only one actor in the final stage being correctly identified by all participants.

One potential explanation for the lack of differences between the two experimental groups is that the experimental learning group did not receive enough training, with alternatively the control group receiving too much prior training. However, both groups experienced the same initial instructions, the learning group only knowing that they were to be attending more than one session. This group was also provided with extensive on using the video facilities to help them make their decisions. There were also no ceiling or floor effects, meaning that task complexity cannot be an explanation. Furthermore, as noted, the performance of the learning group improved throughout the course of the experiment, meaning that continual exposure to the faces was enhancing identification. It was in the final trial only that performance was equivalent across the two experimental groups. It is possible that if the final task had been more complex, perhaps involving the use of arrays of faces instead of the more forensically-valid two alternative-choice design, differences between the groups would have been uncovered. This could be investigated in further experiments, which might therefore provide further information concerning the development of face familiarity.

However, in terms of the applied nature of the current research, pilot participants, highly familiar with the actors in the same videos named the actors 100% correctly, in most cases after viewing the same footage for a few seconds (Chapter 3.5). In contrast, although performance improved throughout all three critical stages, final accuracy in the learning group in this experiment was well below 100%. Implied in the Court of Appeal ruling in *R v Clare and Peach* (1995) is that comprehensive inspection of images as carried out in this experiment should result in full familiarisation to the depicted individuals. From these results it is apparent that participants were unable to sufficiently familiarise themselves with the actors using this procedure, questioning the reliability of this type of evidence.

One further issue when attempting to simulate police procedures is that the focus for a police officer would be a single investigation. However, to ensure enough statistical power to test assumptions, a large number of participants would be required if a single film of one actor was used as a stimulus in this type of study. In addition, the use of a single actor might produce results that do not generalise to the wider population. It could therefore be argued that the results of Experiment 8.1 have limited relevance to forensic scenarios. However, all participants were provided with performance feedback in the final stage of the experiment, which involved reviewing the pairs of images and confirming the accuracy of their written selections. Although no direct measures were taken, many, especially in the learning group, were surprised at the number of errors they made. Most appeared to believe they would be informed that they were extremely accurate, although some admitted that they had not been fully confident in all trials. However, a few were convinced that the feedback was wrong and asked to view the videos and photographic images again in specific trials. In all cases, these highly confident participants were incorrect.

The results of this experiment suggest that caution should be taken when testimony is provided by a witness based on familiarisation to an individual who was previously unknown. However, the global increase in CCTV surveillance is likely to be accompanied by an increase in the number of police officers giving identification evidence of the type described in *R v Clare and Peach* (1995), particularly in the case of minor crimes being prosecuted in a magistrate's court. Therefore, operational procedures may need to be established, perhaps with the use of standardised line ups as normally required of eyewitnesses. To increase the reliability of evidence, any beneficial methodology should decrease the likelihood of false positives while stabilizing the number of false negatives.

## Chapter 9: Jury decision making when presented with CCTV evidence

### 9. 0: Introduction

The experiments reported so far in this thesis have only analysed the individual responses of participants. In a courtroom, members of a jury may be required to jointly provide a verdict partly based on their perception as to whether a culprit shown on surveillance video is the defendant in the dock. Indeed, convictions have been secured on this basis (e.g., *Church v HMA*, 1996; R v *Dodson and Williams*, 1984). Even if they have not specifically been asked to compare images with the accused, jurors can always make their own decisions and these may be based on the resemblance. Judges in the UK normally warn juries of the potential for photographic or CCTV evidence to be misleading (e.g., Attorney General's Reference, No. 2 of 2002, 2003; *R v Dodson and Williams*, 1984). However, the ability to 'see for themselves' may outweigh such cautions.

Although the opinions of individual jurors may not be shared by the remainder, research into jury decision making has found that if at the start of deliberation the majority of a jury believes that the defendant is shown in CCTV, a guilty verdict will probably be rendered. Indeed, in 90% of real trials, a majority initially voting for either a conviction or an acquittal was found to result in a corresponding final verdict (Kalven & Zeisel, 1966; Sandys & Dillehay, 1995). This finding has been replicated in the laboratory using experimental simulated mock juries (e.g., Devine et al., 2001; MacCoun & Kerr, 1988).

However, jury decision making has also been found to conform to statistical models of group decision making, based on the probability of a specific outcome rather than just a numerical majority (Davis et al., 1975; Devine et al., 2001). There are thirteen possible primary voting patterns that may be observed in any twelve person jury (12-0, 11-1, 10-2 etc.) and conformity to the models can be assessed by generating a projected verdict distribution based on these patterns and determining the closeness of fit to the actual data. Early research generally found that a two-thirds pre-deliberation majority in either direction would result in a final verdict in the same direction (e.g., Davis et al., 1975; MacCoun & Kerr, 1988). However, Devine et al. (2001) reviewed 26 studies, across which 348 twelve-person mock juries had been conducted. They found that if hung juries were discounted, due to their comparative rarity in real trials, there was an asymmetrical leniency bias. As such, if ten or more jurors initially vote guilty, conviction will normally result, whereas an acquittal will probably occur if an initial pro-conviction position is shared by seven or less. However, the model cannot account for intermediate preferences (9-3 or 8-4) and the outcome could be in either direction.

These findings are perhaps discouraging, suggesting that regardless of minority concerns the deliberation process is a foregone conclusion. However, criticism has been directed at the use of mock juries for their lack of realism and external validity (e.g., Bornstein, 1999; Bray & Kerr, 1979). Indeed, examination of the voting patterns of initially evenly-split *real* juries (6-6) has found a far weaker leniency effect. For instance, Kalven and Zeisel (1966) found that across ten initially evenly-divided trials, conviction rates were 50%. An even higher conviction rate of 71% from 24 trials was found by Sandys and Dillehay (1995), although as jurors were retrospectively questioned in both studies, the operation of a memory bias cannot be discounted. Indeed, conclusions are limited as these real trials represent far fewer examples than have been conducted in the laboratory. However, few researchers have been allowed to observe or record real jury deliberations in the USA; and in the UK post-trial discussions are illegal.

Particular criticism has been directed at the lack of binding consequences when research is based on simulated juries. Indeed, in a field setting, Diamond and Zeisel (1974) recruited participants in groups to attend ten different real cases as spectators and to act and to privately deliberate as though they were giving a binding verdict. Whereas the real juries convicted the defendants in 50% of cases, the conviction rate from the simulated juries was 90%, illustrating that even in identical circumstances, consequential factors may influence verdicts. In addition, studies utilising a student disciplinary scenario to which a far higher number of participants were recruited have uncovered conflicting or null effects when comparing judgements made knowing decisions were not binding, to those made study-blind (Kerr, Nerenz & Herrick, 1979; Wilson & Donnerstein, 1977). In one, students described as believing that the consequences of their decisions would have a binding effect produced more guilty verdicts than those who knew the case was hypothetical (Wilson & Donnerstein, 1977). In contrast, a second similar study found no differences in verdicts regardless of whether students were study-blind or not (Kerr et al., 1979). However, consequences of decisions were far less serious than would be encountered by most criminal juries.

Even taking into account these conflicting findings, research using mock juries can effectively control for potential extraneous variables while allowing a small number of focal variables to be examined (Devine et al., 2001). In terms of the topic of this thesis, it provides the ideal scenario for examining jury decision making when confronted with surveillance footage. In the UK, during summing up by a judge, juries should be warned, that if they have any reasonable doubt as to the guilt of the defendant, an acquittal verdict should be agreed. Therefore, the aim of the experiment reported in this chapter was to use the mock jury paradigm to examine how juries would respond when presented with CCTV evidence for identification purposes, especially when reminded of this standard of proof.

Also of interest was to compare individual and group responses to investigate whether individual jurors, if initially in disagreement, would polarise their own beliefs to conform to a group norm, or whether they would publicly agree with the majority while privately responding differently. A considerable body of research has been published on group polarisation effects (e.g., Isenberg, 1986; Myers, 1978; Myers & Lamm, 1976; Myers, Bruggink, Kersting & Schlosser, 1980; Sanders & Baron, 1977). In particular, investigations have been conducted on why in some circumstances a group may come to a more extreme or risky final decision than the initial position of its individual members, instead of a more cautious compromise based on the average position of the group.

Two associated mechanisms have been proposed for this effect (Isenberg, 1986; Myers & Lamm, 1976). The *social comparison* model suggests that people prefer to project themselves in a socially favorable light, by appearing distinct and 'individual'. If from the removal of 'pluralistic ignorance' an individual discovers that their own view differs from the perceived central tendency of the group, they will often attempt to shape a compromise, without entirely rejecting their own position. If a majority of group members shift in this manner, a group polarisation effect will occur. This will be enhanced, causing a *risky shift*, if the central tendency is perceived to be more extreme than it is in actuality. This type of effect has been experimentally demonstrated many times, in that 'mere-exposure' to knowledge of the mean opinion of a group can alter the position of an individual within that group, without hearing any actual arguments in favour of the group's opinion (Myers, 1978; Myers et al., 1980, Sanders & Baron, 1977).

An associated *informational influence* mediating explanation for group polarisation is that discussion will generate a series of opinions, most of which will be supported by the majority and therefore will have already been considered by most individuals in advance. However, some arguments will be novel. The degree of opinion shift will be determined by the proportion of arguments supporting one side as opposed to another, as well as their strength of logic, with novel powerful arguments likely to be the *most* persuasive, resulting in extreme shifts of opinion (Bray & Kerr, 1979; Isenberg, 1986). In support, studies have measured a strong correlation between the proportion of persuasive arguments in one direction and the degree of shift towards that position (Ebbesen & Bowers, 1974; Madsen, 1978). Furthermore, a greater reliance on newer arguments over those discussed or considered earlier has been found to have a causal effect on polarisation in a risky direction (Kaplan & Miller, 1977; Vinokur & Burnstein, 1978).

In terms of jury decision making, a significant risky shift would involve a large minority during deliberations moving from initial individual not guilty positions to a final group guilty verdict. A cautious approach would result in an opposite effect. However, most mock jury investigations have found an overwhelming polarisation effect towards leniency (e.g., Davis et al., 1975; Kaplan & Miller, 1977; Kerr, Nerenz & Herrick, 1979), possibly due to the typically more liberal attitudes of the student sample used in this type of research (Bray & Kerr, 1979). Exceptions have tended to come from manipulating jury membership, so that if for example, the majority measure high in authoritarianism, the final verdict will tend to be shift towards greater punishment severity, whereas with the same materials, juries scoring low on the same trait will shift towards leniency (e.g., Bray & Noble, 1978).

**Experiment 9.1: Jury decision making when confronted by CCTV evidence**

The primary aim of the experiment reported in this chapter was to examine simulated group jury decision making when presented with CCTV evidence. To minimise the influence of extraneous factors the participants were informed that the 'guilt' of the defendant was based entirely on whether they were depicted in video footage. Each jury deliberated as a group and was required to conduct a series of polls. They were reminded prior to a final vote that if they had any reasonable doubts they should enter a not guilty verdict. However, a problem with examining jury level decisions is that the requirement of 12 jurors in each trial and the low variability in final response outcomes necessitates large numbers of participants for appropriate statistical analyses. Therefore, data was also assessed at an individual juror level, with participants responding privately before and after deliberation.

In the UK, most jurors are called for a two week period of jury service, often being selected for a number of cases with many of the same cohort. In some states in the USA, jury service can be longer and in Kentucky it is at least one month (Dillehay & Nietzel, 1985). Therefore, the same jurors may plausibly be presented with CCTV evidence in more than one case. Prior jury experience has been found to influence later verdicts. For instance, Dillehay and Nietzel (1985) examining the composition of 175 real trial juries, found that an increase in the proportion of jury-experienced jurors was associated with more guilty verdicts. Mock jury studies have also found that experience influences verdicts. In one (Nagao & Davis, 1980), simulated juries produced verdicts on two consecutive cases. If a rape case was second, juries were less likely to convict than if it had been presented first. The opposite effect was found with a vandalism case. In contrast, Kerr et al. (1982) found no effects of experience in a study in which student participants acted as jurors in 5 different consecutive hypothetical trials.

Although the results of these studies are mixed, the findings that prior experience impacts on jury verdicts has an important implication in terms of the presumption of innocence as well as the burden of proof, particularly the standard of reasonable doubt. Therefore, a further aim of this experiment was to examine whether the requirement to make a second jury decision in which CCTV evidence is crucial to the case, was also partly influenced by previous group and individual verdicts.

Each 'jury' produced verdicts in two 'trials', one in which the target 'defendant' depicted in a photograph was present in video, one in which they were absent. Presentation order was counterbalanced and to increase the likelihood of discussion, the video and photograph selected for each 'case' had been identified as being of the same person in approximately 70% of trials in previous experiments. As changes from participants' original positions must be due to deliberation and the influence of the group it was possible to determine whether they conformed to statistical models of decision making (Davis et al., 1975; Devine et al., 2001).

It was also possible to compare participants' private pre-deliberation and post-deliberation judgements across final jury-level verdicts (e.g., guilty, hung, not guilty) to see if individual juror responses tended to follow theoretical group polarisation models (e.g., Bray & Kerr, 1979; Isenberg, 1986; Myers & Lamm, 1976). These models would specifically predict that individual belief structures as measured by scale responses would be directly influenced by group decisions.

However, the design did not allow a direct assessment of the social comparison or informational influence models of social polarisation.

Instructions were provided to encourage an agreed group verdict. These interventions were based on past research into deliberation processes, reportedly producing faster agreements and less biased decisions (Davis, Stasson, Ono & Zimmerman, 1988; Devine et al., 2001; Hastie, Penrod, & Pennington, 1983; Kerr, 1982). For instance, Hastie et al. (1983) identified two approaches adopted by juries. Once a foreman is selected, 28% initiate an immediate poll with deliberations tending to be *verdict-driven*. A further 35%, *evidence-driven* juries, postpone voting until after comprehensive discussion with the remainder using a combination of both. Davis et al. (1988) suggest that the likelihood of an individual altering their voting preference is positively correlated with an increased regularity of polling and Kerr (1982) found that regular voting is associated with fewer hung verdicts. In addition, voting system mechanics can affect outcome. For instance, if they adopted a sequential system of voting, Davis et al. (1988) found that when a not guilty faction voted first, juries were 14% more likely to acquit (75%), than if a guilty faction voted first (61%). There was an intermediate outcome when juries conducted simultaneous votes (69%). Therefore, the juries in this experiment were encouraged to use a verdict-driven approach by conducting an immediate poll and to vote as often as possible thereafter. Furthermore, to reduce the likelihood of any voting mechanic bias, foremen were instructed to conduct simultaneous polls.

Finally, unanimous jury decisions are normally encouraged in the UK, partly to eliminate one potential justification for an appeal. However, if less-than-unanimous majority decisions (e.g., 10-2; 9-3 etc.) are allowed, final verdicts in mock jury trials rarely differ from those in which unanimous decision are required (Davis et al., 1975; Devine et al., 2001; Nemeth, 1977). Davis et al. (1975) also found that majority decisions are associated with reduced deliberation time, less hung decisions and fewer polls, than unanimous decisions. Therefore, for the purposes of analyses, majority decisions of 10-2 and above were regarded as definite verdicts, with any other outcome regarded as a hung decision.

### 9. 1. 1. Method

#### 9. 1. 1. 1. Participants

Participants were 101 female and 21 male first year adult undergraduate students at Goldsmiths College, University of London (Mean age = 21.9, *SD* = 6.4). All gained course credit for participation. None of the participants had taken part in any of the previous experiments or pilot studies.

#### 9. 1. 1. 2. Materials

Two of the 30-sec medium-range videos depicting different actors in no disguise were used (Chapter 3.3.2). A three-quarter close-up A4 sized black-and-white facial photograph of one of the two videoed actors (the 'defendant') was used in target present conditions. A similar sized photograph of a matched distracter was used in target absent trials. Stills from the videos, and the photographs used are depicted in Figure 9.1. These pairs of images had also been used in the single-item identity-verification experiments reported in Chapters 4 and 5. Across all the trials in those experiments the positive identification rate for these stimuli was the same, even though one condition was target present, the other target absent. Indeed, in the target present trials in which Actor 08 was depicted in both photograph and video footage, 63 out of 174 (36.2%) participants *incorrectly* responded that two different people were shown. In target absent conditions when Actor 36 was depicted in video and Actor 37 was in the photograph, 63 out of 174 (36.2%) participants *correctly* responded that two different people were depicted.

For the measurement of juror-level individual private responses, two 8-point identity-belief scales were provided to each participant, one prior to deliberation, and the other post-deliberation. The form containing the post-deliberation scale allowed space for participants to note any comments about the procedure and to identify influential jurors. For the recording of juror-level public voting, the foreman of each jury was given a response sheet upon which there was space to document up to twelve polls. Participants were allocated an identifying letter (A – L) logically based on seating position to allow the foreman to keep a voting record.

*Figure 9.1: Clockwise from top left, Actor 08 depicted in two still images from video and in a photograph for target present trials. Actor 36 depicted in photograph and two video stills with Actor 37 in a photograph (below left) for target absent trials.*

**9. 1. 1. 3. Design**

Participants were required to act the part of individual jurors and to make a series of decisions as to the 'guilt' or not of photographed 'defendants' during two separate simulated 'trials' in which the evidence presented consisted of a single security video allegedly depicting that defendant. Verdicts were rendered individually (juror-level) and as a group (jury-level). Jury-level examination was conducted on verdict outcome based on whether the defendant was actually present or absent in the video. The primary juror-level design was a 2 (target presence) x 2 (pre/post-deliberation) x 2 (trial order) mixed design. The first repeated-measures factor was target presence, one trial was target present and one was target absent. The second repeated measures factor was that jurors returned two private responses using the identity-decision scale during the course of each trial; one pre-deliberation and one post-deliberation. The independent-measures variable was that the trial order was counterbalanced so that half the participants took part in the target present trial first. The remaining participants took part in the target absent trial first. The scale used allowed analyses of unconverted scores, verdicts and confidence as separate dependent variables.

**9. 1. 1. 4. Procedure**

Participants initially signed up to participate in groups of 13, to allow for non-attendance. They were seated in a U-shaped curve facing a 26" colour television/video screen. Each participant was provided with an identifying letter and the foreman was randomly selected based on seating position. If there was no participant drop out, the foreman would contribute individual data to the experiment, but during the deliberation phase would act as a chairperson and would not vote in any poll. Otherwise, the foreman acted as a normal member of the jury. After being given instructions, participants were required to view one of the 30-sec simulated surveillance videos depicting the 'culprit' and compare it to a simultaneously presented 'defendant' in a photograph located alongside the television screen. The experimenter enforced silence throughout this stage and the participants indicated their private belief as to identity using one of the identity-decision scales. There was no time limit, but in most cases the video was shown five times.

The experimenter collected these forms and left the room after verbally reminding participants of the written instructions for conducting their deliberations. The foreman was immediately required to carry out the following public vote procedure. Firstly, participants simultaneously raised their hands if they believed the person in the video was depicted in the photograph ('guilty'), followed secondly by polls of those believing the opposite ('not guilty') and finally of those abstaining. They were instructed to discuss the case in an attempt to arrive at a unanimous verdict and to re-vote as many times as required. It was emphasised that they should treat the procedure as seriously as if the fate of the defendant rested upon their combined decision.

After 15-min the experimenter re-entered the room and informed the jury that if a unanimous decision had not been reached they should attempt to reach a 10/2 majority. After another 5-min, the experimenter again re-entered the room and asked the foreman to conduct a final vote, specifically asking the jury to attempt to "come to a unanimous decision as that is what is normally expected in a court of law. In a real jury, if you have reasonable doubts as to the guilt of a suspect you would be directed to vote not guilty" and that abstentions were prohibited. Following this poll, participants privately completed the second identity-decision scale. They were told that this decision should be "completely independent of any previous decision they had made, including the vote they had just been involved in". They were then asked to make comments about the procedure and points made by individual jurors on the response sheet provided. Response forms were collected and the above procedure was repeated using the counterbalanced target presence stimuli. Participants were then fully debriefed.

### 9. 1. 2. Results

In two of the ten jury groups, 13 participants contributed data and in these the foreman was instructed not to publicly vote. However, when appropriate, their individual private responses are analysed below. In 12-person juries, all participants contributed data to all stages of the experiment.

#### 9. 1. 2. 1. Jury-level public decisions

Each jury participated in two counterbalanced 'trials', one target present and one target absent. During the deliberation process, the foreman was encouraged to conduct a series of public polls. The maximum number of votes in any trial was 5, the minimum 2. A repeated-measures t-test found that foremen conducted marginally more polls in first trials ($M = 3.9$, $SD = 0.74$) than in second trials ($M = 3.2$, $SD = 0.63$; $t(9) = 1.91$, $p = .091$. There were also more polls in target present trials ($M = 3.8$, $SD = 0.63$) than in target absent trials ($M = 3.3$, $SD = 0.82$). However, a similar t-test found that this difference was not significant, $t(9) = 1.25$, $p > .1$.

Apart from during the final poll, participants responded with guilty, not guilty or unsure verdicts. The final poll required a guilty or not guilty vote. Table 9.1 lists the final verdict outcomes for all 20 trials as a function of initial preferences dependent on whether trials were target present (TP) or absent (TA). The numbers in the body of the table refer to the number of target present and target absent trials in each verdict category. For the purpose of display, unsure decisions are treated as not guilty preferences and a final majority of more than 10-2 in either direction is treated as a guilty or a not guilty verdict.

*Table 9.1: Summary of predeliberation verdict distributions and final jury verdict outcomes as a function of target presence*

| Initial preference distribution | Final verdict frequencies | | |
|---|---|---|---|
| G, NG | Guilty | Hung | Not Guilty |
| 11,1 | 1TP | - | - |
| 10,2 | - | - | - |
| 9,3 | 1TA | - | - |
| 8,4 | 1TP | - | - |
| 7,5 | - | 1TP | - |
| 6,6 | - | 2TP 1TA | 1TA |
| 5,7 | - | - | 1TP |
| 4,8 | - | 1TA | 1TA |
| 3,9 | - | 1TA | 1TP 2TA |
| 2,10 | - | 1TP | 2TP 2TA |

*Note: TP = Target present trial; TA = Target absent trial; G = Guilty; NG = Not Guilty*

Of the twenty trials, four resulted in unanimous decisions (12-0), one of which was a target present trial resulting in a correct guilty verdict. The remaining three trails in which a unanimous verdict was reached (two target present; one target absent), resulted in not guilty verdicts. A further nine trials resulted in majority verdicts of at least 10-2. In one of these, the final majority wrongly voted guilty when in fact the photograph and video depicted two different people. The remainder had evenly-split final verdicts or majorities of less than 10-2. These were classed as hung verdicts.

**9. 1. 2. 2. Correspondence between private and public decisions**

Participants completed one of the individual 8-point identity belief scales immediately preceding the first public poll. They also completed a second immediately after the final public poll. It was therefore possible to examine the correspondence between these responses, as, for instance, a high private scale score (5 – 8) would be expected to be matched with a publicly made individual guilty vote.

Ten out of the total of 240 (4.2%) individual pre-deliberation scale responses directly contradicted those participants' initial public voting choices and more (20/240; 8.3%) post-deliberation scale responses differed from those given in the final public group poll. In the majority of the post-deliberation cases ($n$ = 12) the final group jury verdict was a not guilty decision, indicating that despite *publicly* agreeing with the majority, these participants *privately* believed that the defendant was depicted in video. The remainder of the final incongruent individual private/public decisions occurred in trials in which the final verdict was hung. No participant publicly responded with a 'guilty' verdict while privately reporting that the defendant was absent from the footage.

To measure correspondence between public voting and private responses, public guilty votes were given the value of 2; unsure votes 1; and not guilty votes 0 and two sets of four Spearman's correlational coefficient analyses were conducted. The first set was between private identity-scale unconverted scale scores and converted public vote scores. For the second set the scale scores were converted so that low scale scores (1 – 4) were treated as not guilty responses and given the value of 0 and high scale scores (5 – 8) as guilty responses and given the value of 1.

These analyses revealed a high correlation between initial public votes and predeliberation (*i*) unconverted 8-point scale scores, and (*ii*) converted 'guilty or not' scale scores in both target present, $r_i(120) = .819, p < .001$; $r_{ii}(120) = .855, p < .001$ and target absent trials, $r_i(120) = .779, p < .001$; $r_{ii}(120) = .876, p < .001$.

There was also a high correlation between final public votes and post-deliberation (*i*) unconverted scale scores, and (*ii*) converted 'guilty or not' scale scores in target present, $r_i(120) = .846, p < .001$; $r_{ii}(120) = .874, p < .001$ and target absent trials, $r_i(120) = .720, p < .001$; $r_{ii}(120) = .827, p < .001$. These high correlations indicate that the individual private scale responses corresponded closely to the majority of jurors' public decisions. Therefore, further analyses were conducted on the private scale scores as they allow for greater analytical sensitivity.

#### 9. 1. 2. 3. Juror-level private decisions

Using the converted scale scores, regardless of which trial participants took part in first, 46.8% of target absent pre-deliberation decisions were incorrect, whereas 32.8% were incorrect post-deliberation. In target present trials, 37.8% of pre-deliberation decisions were incorrect whereas 50.8% were incorrect post-deliberation. A series of three-way ANOVAs were conducted to examine the effects of trial order and deliberation in target present and target absent trials on the unconverted scale scores, verdicts and confidence level data.

#### 9. 1. 2. 3. 1. Juror-level private unconverted scale scores

A 2 (target presence: present vs. absent) x 2 (deliberation: pre- vs. post-deliberation) x 2 (trial order: first trial experienced vs. second trial experienced) mixed design ANOVA was conducted on the juror-level private unconverted scale score data. Figures 9.2a and 9.2b shows the mean pre- and post-deliberation scale scores dependent on whether participants experienced the target present trial first or the target absent trial first. To present these figures in terms of the 3-way ANOVA conducted on these data, it would be customary for the target present values in Figure 9.2b to be presented to the left, with the target absent values presented on the right of the figure. However, to increase the clarity of the data, it is presented in terms of the *order* of trial type experienced by participants in each counterbalanced condition, with the first type of target presence trial presented on the left of the figure.

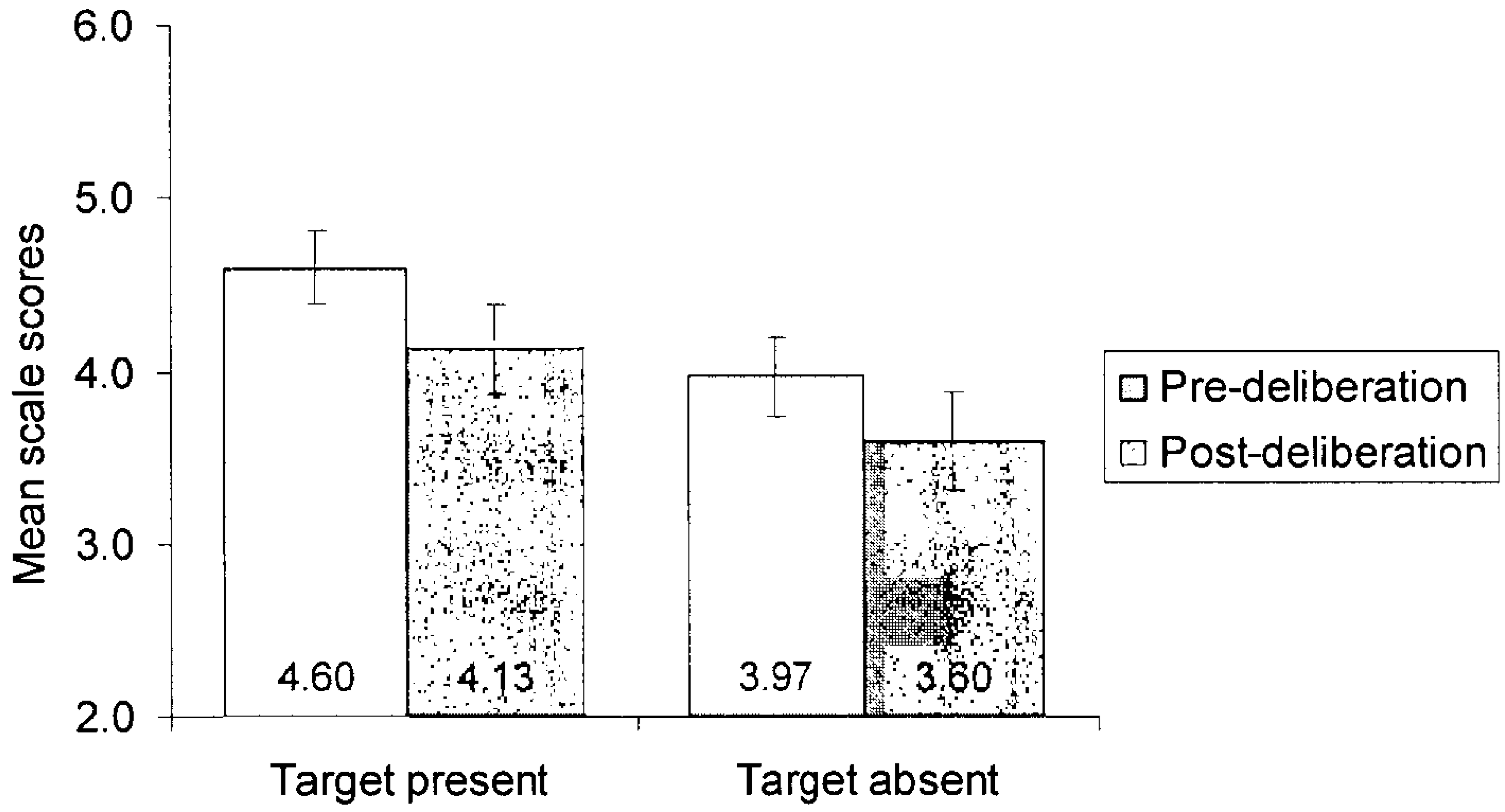



*Figure 9.2a: Private identity-scale responses (maximum = 8) for participants who experienced the target present trial first (error bars denote standard error of the mean)*

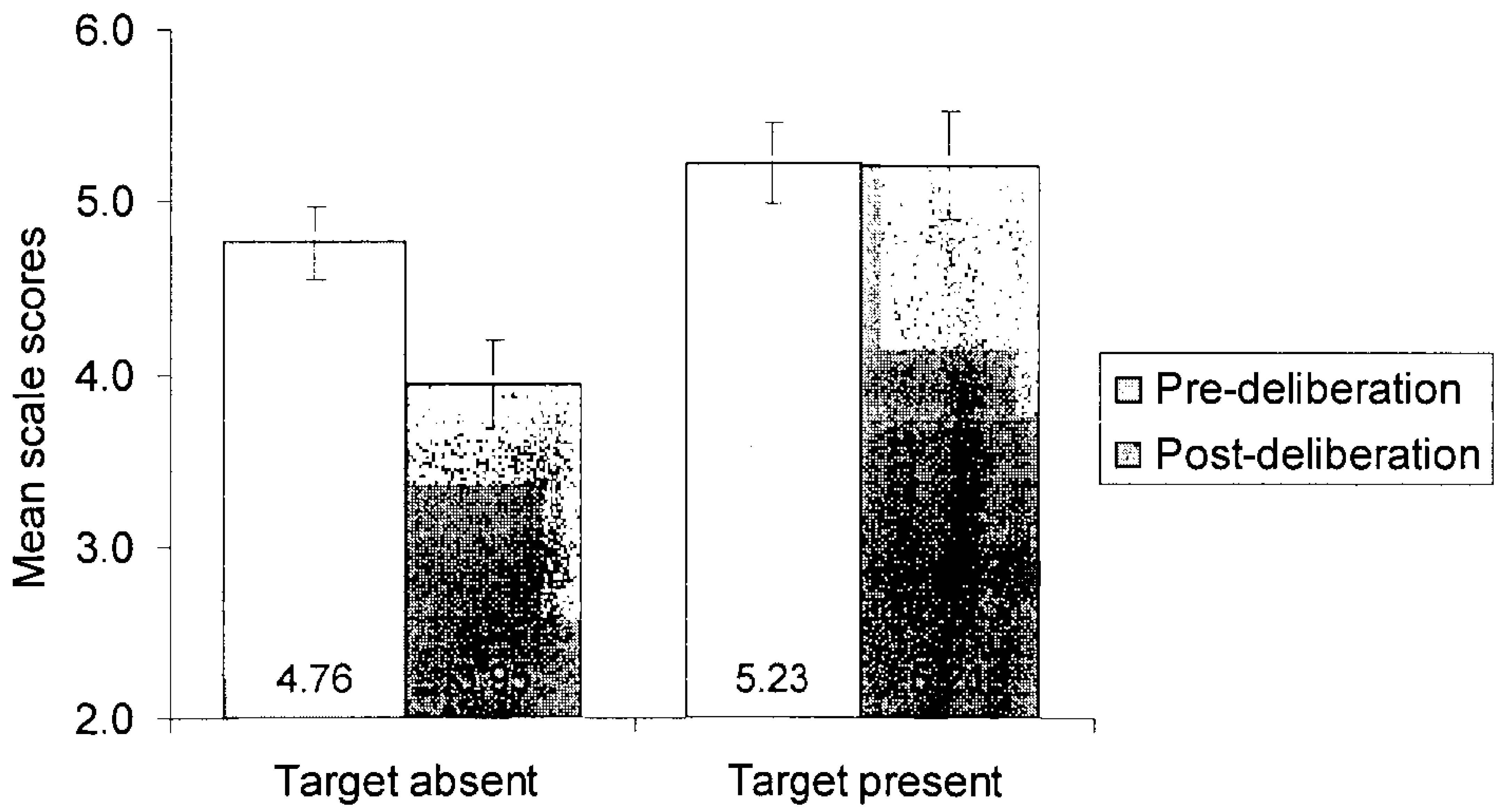



*Figure 9.2b: Private identity-scale responses for participants who experienced the target absent trial first (error bars denote standard error of the mean)*

The repeated measures main effect of target presence was significant, $F(1, 120) = 9.09, p < .005$; scores were higher when targets were present than when absent ($M = 4.80, SD = 1.90$ vs. $M = 4.07, SD = 1.79$). The repeated measures main effect of deliberation was also significant, $F(1, 120) = 17.32, p < .001$; pre-deliberation scores were higher than post-deliberation scores ($M = 4.64, SD = 1.30$ vs. $M = 4.23, SD = 1.50$). The independent measures main effect of decision order was also significant, $F(1, 120) = 9.87, p < .005$; scores were higher when participants took part in target absent trials first than when they took part in target present trials first ($M = 4.79, SD = 1.32$ vs. $M = 4.07, SD = 1.18$).

There was a trend towards a significant interaction between target presence and deliberation, $F(1, 120) = 3.14, p = .079$. However, this was mediated by a significant three way interaction, $F(1, 120) = 5.22, p < .05$. Bonferonni-corrected simple interaction effects found that when the target present trial was conducted first, only the effect of deliberation was significant, $F(1, 59) = 9.37, p < .01$. As such, pre-deliberation scores were higher than post-deliberation scores as reported for the main effects above. The effect of target presence and the interaction were not significant ($p > .1$). However, when the target absent trial was conducted first, the effects of deliberation, $F(1, 61) = 8.07, p < .05$; and target presence $F(1, 61) = 6.64, p < .05$ were both significant and consistent with the main effects reported above. However, the interaction was also significant, $F(1, 61) = 8.20, p < .05$. Repeated-measures t-tests revealed that in the initial target absent trial, scores were reduced following deliberation, $t(61) = 3.88, p < .05$. In contrast, in the subsequent target present trial, scores did not significantly alter ($p > .1$).

#### 9. 1. 2. 3. 2. Juror-level private verdicts

The unconverted scale data indicates whether the responses of participants changed. However, it does not show whether the actual verdict preferences were altered following the deliberation process. Therefore, individual private scale responses were converted to produce verdict data by giving scores of 5 – 8 a value of 1 (guilty) and scores of 1 – 4 a value of 0 (not guilty). Figures 9.3a and 9.3b show the mean verdict data pre- and post-deliberation dependent on whether participants experienced the target present trial first or the target absent trial first. As in Section 9.1.2.3.1, these data are presented in terms of the order in which participants experienced their target presence trials, with the type of trial experienced first presented on the left of the figure.

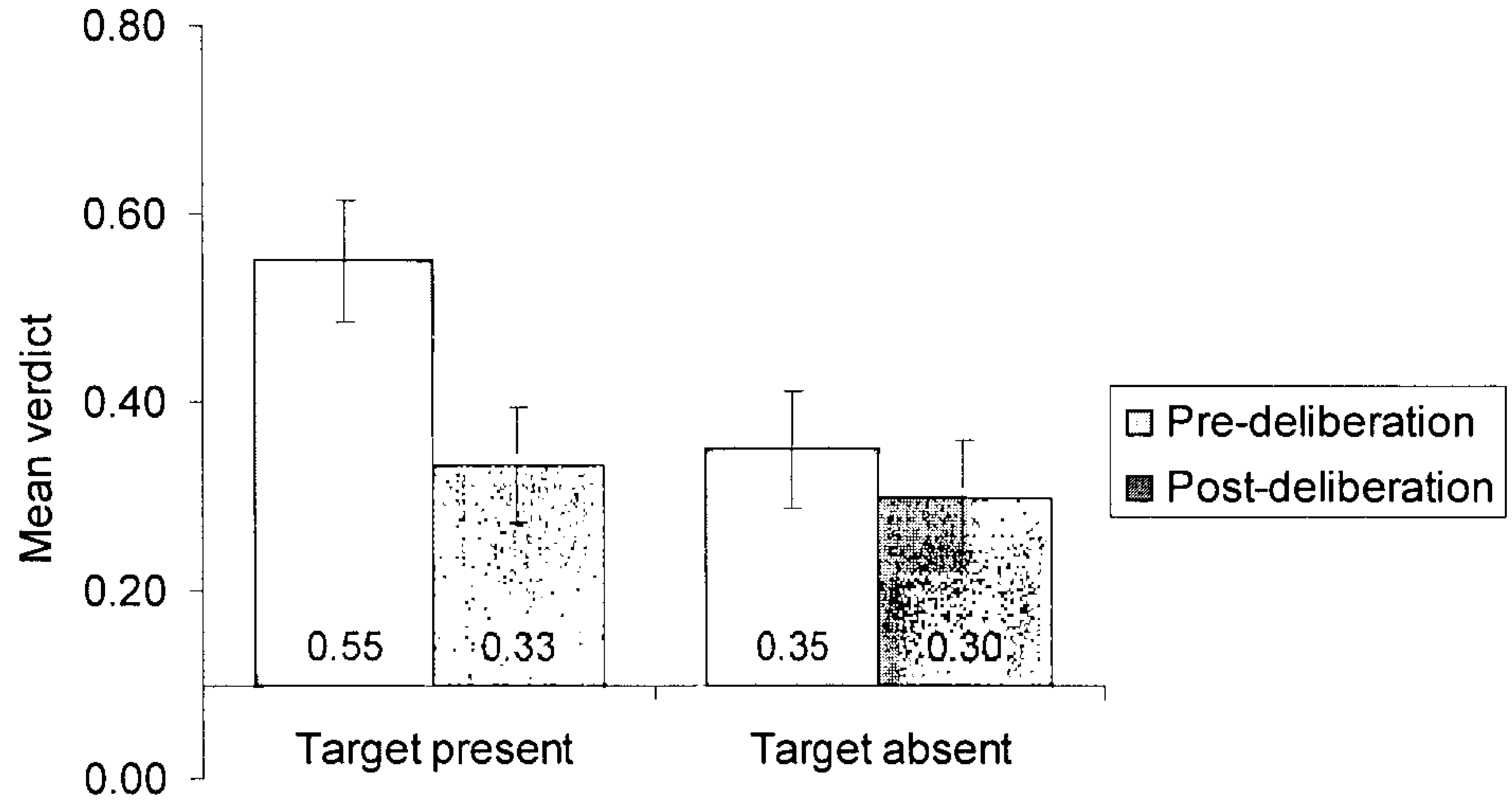



*Figure 9.3a: Mean verdicts (1 = guilty; 0 = not guilty) when participants experienced the target present trial first (error bars denote standard error of the mean)*

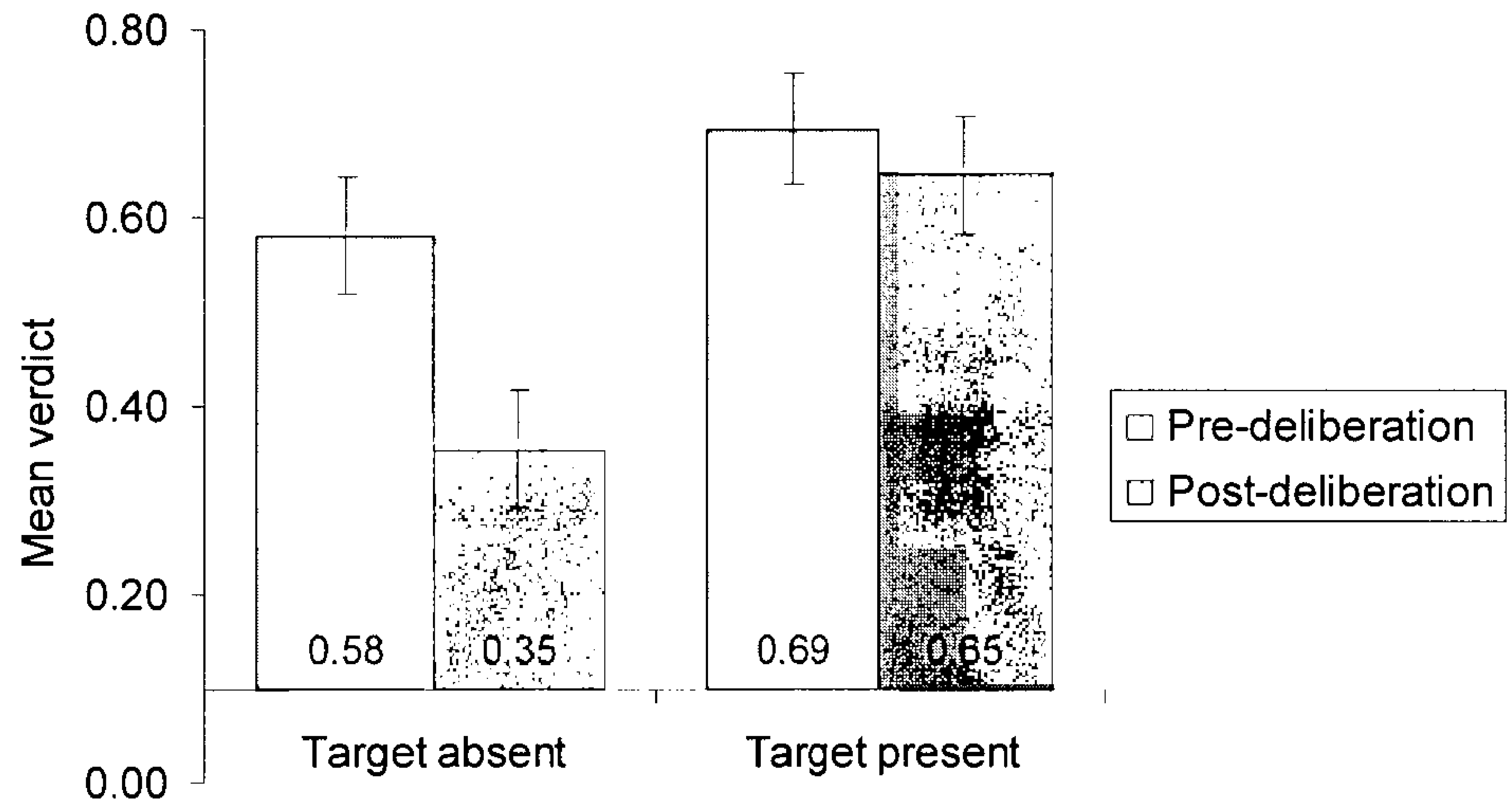



*Figure 9.3b: Mean verdicts (1 = guilty; 0 = not guilty) when participants experienced the target absent trial first (error bars denote standard error of the mean)*

A 2 (target presence) x 2 (deliberation) x 2 (trial order) mixed design ANOVA was conducted on these data. The main effect of target presence was significant,

$F(1,120) = 8.74$, $p < .005$. Participants were more likely to vote guilty in target present conditions than in target absent conditions. The main effect of deliberation was also significant, $F(1,120) = 18.43$, $p < .001$. Participants were more likely to vote guilty prior to deliberation than post-deliberation. The main effect of trial order was also significant, $F(120) = 11.87$, $p < .001$. Participants were more likely to vote guilty when the target absent trial was experienced first than when the target present trial was experienced first. None of the two-way interactions were significant ($F < 1$ for all).

However, the three-way interaction was significant, $F(1, 120) = 9.14$, $p < .005$. Bonferonni-corrected simple interaction effects found that when the target present trial was conducted first, the effect of deliberation was significant, $F(1, 61) = 8.47$, $p < .01$. Participants moved from a guilty to a not guilty verdict preference following deliberation in both trials. The effect of deliberation and the interaction were not significant ($p > .1$). When the target absent trial was conducted first, the effects of target presence, $F(1, 61) = 7.08$, $p < .05$; deliberation, $F(1, 61) = 10.03$, $p < .01$ and the interaction, $F(1, 61) = 5.16$, $p < .05$ were all significant. Participants were more likely to move towards a not guilty verdict following deliberation but only in the first target absent trial. Verdicts did not change in the second trial when targets were present.

#### 9. 1. 2. 3. 3. Juror-level private confidence

Individual scale scores were also converted for confidence and a 2 (target presence) x 2 (deliberation) x 2 (trial order) mixed design ANOVA was conducted on this data. The independent-measures main effects of target presence, $F < 1$, and trial order, $F < 1$ were not significant. However, the repeated-measures main effect of deliberation was highly significant, $F(1, 120) = 44.02$, $p < .001$. Confidence was higher following deliberation than prior to deliberation ($M = 2.48$, $SD = 0.75$ vs. $M = 2.07$, $SD = 0.68$).

The interaction between target presence and trial order was also significant, $F(1, 120) = 9.15$, $p < .005$. Bonferonni-corrected repeated-measures t-tests revealed that when the target present trial was experienced first, there was no difference in confidence between the target present and the target absent trials ($p > .1$). However, when the target absent trial was experienced first, confidence was lower in that trial than in the subsequent target present trial, $t(61) = 2.80$, $p < .05$.

The two-way interactions between deliberation and trial order, $F < 1$; and between deliberation and target presence, $F < 1$ were both non-significant. However, there was a trend towards a significant three-way interaction, $F(1, 120) = 3.22$, $p = .075$. Bonferonni-corrected simple interaction effects revealed that when the target present trial was experienced first, only the effect of deliberation was significant, $F(1, 59) = 21.70$, $p < .001$ as reported for the main effects above. The effect of target presence and the interaction between target presence and deliberation were non-significant ($p > .1$). In contrast, when the target absent trial was experienced first, although the effect of deliberation was again significant as reported above, $F(1, 61) = 23.02$, $p < .001$, the effect of target presence was also significant, $F(1, 61) = 7.84$, $p < .05$. Confidence was higher in the second target present trial than in the first target absent trial. The interaction was between deliberation and target presence was not significant ($p > .1$).

### 9. 1. 2. 4. Group polarisation effect upon individual private responses

It was possible to examine whether there was a shift in individual responses, from their pre-deliberation position to a different post-deliberation position dependent on the final jury verdict, regardless of whether this was correct or not. As there were three verdict outcomes (guilty, not guilty and hung), and participants only took part in two jury-level deliberations, data from the two trials attended by each jury were treated separately. However, pre- and post-deliberation responses were again entered as a repeated-measures factor as was the independent-measures, trial order variable. These analyses were conducted to examine whether the group discussions and voting had altered individual sentiment and not whether decisions were correct or not. Therefore, data from the target present and absent trials were pooled (Figure 9.4).

A 2 (deliberation) x 2 (trial number) x 3 (verdict outcome) mixed design ANOVA was conducted on the individual unconverted scale scores. The main effects of deliberation, $F(1, 238) = 1.08, p > .1$; and of trial number were not significant, $F < 1$. However, the main effect of verdict outcome was highly significant, $F(2, 238) = 32.40, p < .001$; Games-Howell post-hoc tests, conducted due to heterogeneity of variance, revealed that scale responses were highest when the final outcome was a guilty vote and lowest when the outcome was not guilty ($M = 6.44$; $SD = 1.30$ vs. $M = 4.59$; $SD = 1.84$ vs. $M = 3.81$; $SD = 1.64$; $p < .01$ all comparisons).

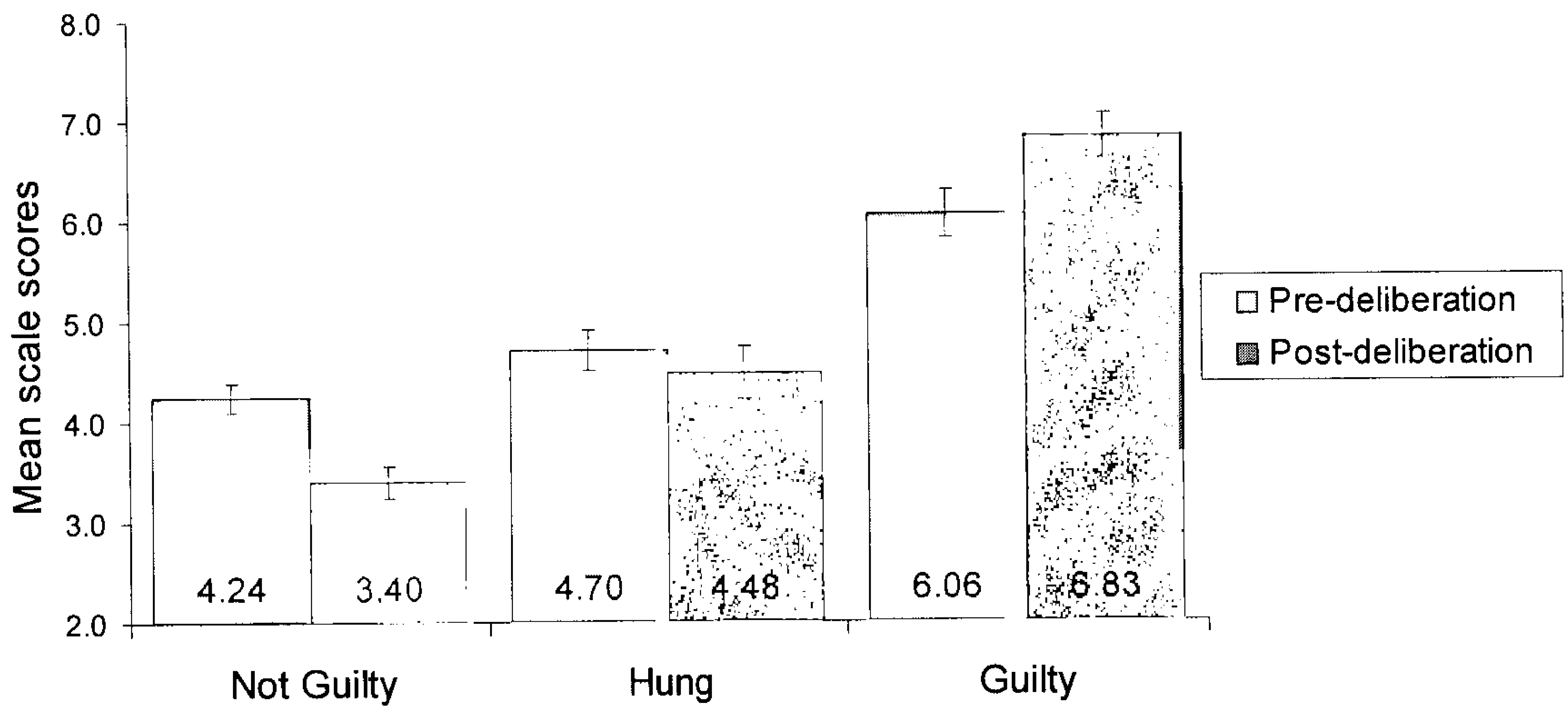



*Figure 9.4: Mean scale scores as a function of deliberation (pre- and post-) and final jury verdict (error bars denote standard error of the mean)*

There was a significant interaction between deliberation and verdict outcome, $F(2, 238) = 15.02, p < .001$. Three repeated measures t-tests were conducted examining the simple effects of deliberation within each level of final verdict. These revealed that scores were significantly higher following deliberation when the final verdict was guilty, $t(35) = -4.59, p < .001$ and significantly lower post-deliberation when the final verdict was not guilty, $t(134) = 6.56, p < .001$. However, there was no significant change when the final verdict was hung, $t(72) = 1.19, p > .1$. None of the other two-way interactions or the three-way interaction was significant, $p > .1$.

Due to low numbers in some cells, similar group-polarisation analyses could not be conducted on accuracy and confidence data.

### 9. 1. 3. Discussion

Experiment 9.1 examined individual and group decision making when presented with simulated CCTV evidence in mock jury trials. Ten 'juries' were formed, each delivering verdicts in two consecutive counterbalanced 'trials', in one of which the photographed 'defendant' was depicted in the surveillance video evidence. In the other, the photograph was of an 'innocent' defendant. Three of the twenty trials resulted in guilty verdicts, one with a unanimous vote, the remainder being 10-2 majorities. All participants were reminded of the beyond reasonable doubt burden of proof and these results suggest that people of jury-age might deliver a guilty verdict based on their perception as to whether someone is depicted or not in surveillance video. Of particular concern was that in one of the three juries rendering a guilty verdict, the 'defendant' was not the depicted 'culprit'. Nine further trials resulted in 10-2 majorities or unanimous acquittal verdicts. Two of these were target present trials and if replicated in a courtroom would represent cases in which a perpetrator escaped justice. As real juries have been invited to compare the defendant with video footage, these findings question the safety of this procedure (e.g., *R v Blenkinsop*, 1995; *R v Dodson and Williams,* 1984; *R v McNamara,* 1996).

The culprit videos and the defendant photographs were specifically selected to induce disagreement amongst participants. In this experiment, prior to any group discussion, approximately 47% of participants in the target absent trials made incorrect responses, with 38% of participants initially incorrect in target present trials. These percentages varied within each jury. From a forensic perspective the final jury-level verdicts are of most interest as they are indicative of what could potentially occur in a real court case, although the limited numbers of trials are insufficient to draw strong conclusions. However, results were consistent with previous studies finding that final verdicts can be predicted by initial juror pre-deliberation preferences (e.g., Davis et al., 1975; Devine et al., 2001; MacCoun & Kerr, 1988). Indeed, there was an asymmetrical leniency bias. All juries that had *less* than 8 pro-conviction pre-deliberation members ended up with either a hung or an acquittal final verdict. All juries with 8 or *more* pro-conviction rendered a final guilty verdict.

Of secondary interest was to examine the influence of jury experience on verdicts, as participants were required to deliberate in two trials. When the group polarisation data is examined in which trials by the same juries were analysed separately, there were no effects of trial number. Regardless of whether it was their first or second trial, juror responses followed a theoretical group polarisation model (e.g., Bray & Kerr, 1979; Isenberg, 1986; Myers & Lamm, 1976), in that individual sentiment post-deliberation was strongly determined by the predominant position, or central tendency prior to deliberation. As no quantitative measures were taken as to the number of arguments discussed during deliberation it was not possible to evaluate the informational influence model of group polarisation. However, the data supported the social comparison model in that if a majority held a pre-deliberation pro-conviction sentiment, and the final verdict matched that position, individual private post-deliberation scores were altered in line with this verdict. The opposite effect was found with juries voting not guilty. No changes were observed with hung juries. As scale score responses were private, these data indicate that individual beliefs in this context can be confirmed or altered by group discussion.

Although the group polarisation data found no trial order effects, mainly because target present and target absent data was combined, the unadjusted scale score and verdict data revealed strong effects of trial order, in which the same data were separated. Scores and verdicts in first trials were equivalent, regardless of whether target present or target absent trials were conducted first, with a reliable movement throughout deliberation towards leniency. A similar pattern was found in the second (target absent) trial for those who experienced the target present trial first. Different effects were found for the group experiencing the target absent trial first. In their second (target present) trial, overall scores were higher than in their first trial and were not changed by deliberation. This finding is consistent with research on real juries finding that increased juror experience is related to an increase in guilty verdicts (e.g., Dillehay & Nietzel, 1985). However, not all studies have found this effect (e.g., Nagao & Davis, 1980; Kerr et al., 1982). These findings, if replicated in a real courtroom would have serious implications, if the procedural aspects of a trial can bias final verdicts.

One of the basic assumptions of the jury system is that deliberation processes will be dominated by those with good memory and confidence of what they have seen in court, not those with higher personal confidence. In this experiment, the process of deliberation was not fully accessed, for instance, by the use of video. However, retrospective written feedback provides an indication of the main discussion topics and the identity of dominant jurors. In many cases reports by jurors in the same trial were contradictory, perhaps illustrating the different importance participants placed on the discussion. Nevertheless, the written feedback is particularly instructive in the single trial rendering an incorrect guilty verdict on an 'innocent defendant'. According to eight jurors, and from a juror's own self report, one claimed expertise from an artistic background into the structure of a face. Similar comments were made in this group's second trial, in which again an incorrect verdict was rendered, although, this time the jury wrongly voted for acquittal. When given performance feedback by the experimenter, this group was particularly surprised by their errors.

If replicated in a courtroom, this finding illustrates that jurors can be strongly persuaded by only one highly influential juror, and the specific 'novel' argument they can provide. Investigations into group polarisation effects have also demonstrated the impact of novel arguments on group decisions (Kaplan & Miller, 1977; Vinokur & Burnstein, 1978), sometimes resulting in extreme shifts of opinion

(Bray & Kerr, 1979; Isenberg, 1986). This particular jury cannot be described as undergoing a 'risky shift' in terms of its final decision, as the majority of jurors tended to incorrectly believe in the 'guilt' of the target prior to discussion. However, these results provide support for the informational influence model of group decision making in that polarisation effects are determined by the number and quality of shared prior opinions as well as the significant influence of newer arguments.

Further criticisms that could be directed at this experiment concerns its validity and in particular the recruitment of an exclusively student sample. Indeed, all mock jury studies can be questioned as to their validity, being artificial in nature and lacking in real binding consequences (Bornstein, 1999; Bray & Kerr, 1979; Kerr et al., 1979; Wilson & Donnerstein, 1977). In addition, in real juries multiple variables may each be weighted by individual jurors differently (Cutler, Penrod & Stuve, 1988), whereas in this experiment there was no prosecution or defense evidence other than the simulated surveillance video. However, all participants in Experiment 9.1 were eligible for jury service, and studies examining this have found only minor effects on mock jury decision making when comparing students to a sample drawn from the wider population (Bornstein, 1999; Cutler et al., 1990).

In addition, the results of this experiment address some of the potential criticisms of the previous experiments in this thesis in which inferences have been made about the potential judgements of real jurors provided with CCTV evidence. There is an implicit assumption concerning jury decision making that the collective intellectual ability of a group is more than just the sum of the individual members. The experiment reported here demonstrated that regardless of the ecological validity of the methodology, participants collectively made errors in judgement as a group based mainly on the proportion making an initially incorrect personal judgement. The group afforded no substantial safeguard against incorrect decisions. Nevertheless, in a real case, a judge in summing up should warn the jury of the potential for error when identification judgements are made to photographic or video evidence. However, as previously noted, members of a jury may interpret this as jargon and as a form of necessary 'legalese', and perhaps disregard it if the majority is convinced by being able to see for themselves any resemblance between video images and the defendant.

Finally, in real court cases, prosecution or defence expert witnesses may be called to provide assistance to juries in order that they may render a correct verdict. In terms of when identification is disputed in CCTV evidence, photographic image analysts from different disciplines may be called in order that they can employ their specialist knowledge to give an indication of whether the defendant is depicted in the footage. Although witnesses providing expert testimony on other types of evidence appear to have a surprisingly small effect on jury verdicts (e.g., Cutler, Penrod & Dexter, 1989; Hosch, Beck & McIntyre, 1980), no published research appears to have examined photo analysts influence on jury decision making. It is unclear therefore whether in these cases, a jury may be more likely to be swayed by the expert testimony. The following chapter describes some of the methods used by expert witnesses in these cases, and in particular evaluates the methodology of one specific technique.

## Chapter 10: Photographic comparison facial individuation techniques

### 10. 0: Introduction

In the UK, if identification from CCTV evidence is disputed, there are four approaches to resolving the issue in court (Attorney General's Reference, No. 2 of 2002, 2003). Firstly, a witness claiming to be familiar with the defendant may give opinion testimony as to whether they believe the defendant is depicted in the images. Secondly, if images are 'sufficiently' clear, the jury may compare them directly with the defendant and be invited to conclude that it is the same person. Thirdly, an 'ad-hoc' expert witness may claim to have familiarised themselves to those depicted based on extensive inspection of the images. The experiments reported in this thesis have demonstrated the potential for an unsafe conviction based entirely on the second and third principles. If no witness familiar with the individual on CCTV is located, a fourth principle can be applied, in that experts in facial structure from a number of different professions can give:

> "Opinion evidence of identification based on a comparison between images from the scene (whether expertly enhanced on not) and a reasonably contemporary photograph of the defendant, provided the images and the photograph are available for the jury" (Attorney General's Reference, No. 2 of 2002, 2003).

The different techniques used by expert witnesses are discussed in this chapter. This is followed by details of an exploratory study into one of these methods designed to provide an estimation of whether two different images depict the same person. For this purpose, a custom software package was created in association with the author of this thesis by Rob Davis, an information technology specialist within the Psychology Department at Goldsmiths, University of London.

### 10. 1: The role of expert witnesses in court

In the UK, it is the prerogative of a judge to determine whether expert witnesses are required, although they should only be called if they can provide the court with "information which is likely to be outside the experience/knowledge of a judge or jury" (*R v Turner*, 1975). The Association of Chief Police Officers (ACPO, 2003) has specified minimum practitioner qualities for facial analyst experts. These include knowledge of facial anatomy, anthropometry, physiology and photographic image analysis techniques. They also acknowledge that dependent on the circumstances of the case different techniques may be necessary, and that "expertise is generally achieved through experience and is measured by the acceptance of reports presented in court" (p. 8). However, the methodology used by photographic comparison experts has not been without criticism and often two different witnesses using similar techniques will come to different conclusions (e.g., *Church v HMA*, 1996; *R v Clarke*, 1995; *R v Gardner*, 2004; *R v Gray*, 2003; *R v Loveridge*, 2001). Moreover, following the later discrediting of the specific methodology used by one particular expert witness, at least one earlier conviction was deemed as unsafe and this was overturned on appeal (e.g., *R v Gray*, 2003).

No published studies appear to have measured the impact of testimony from photographic comparison expert witnesses on jury-decision making. However, parallel research has been conducted on the influence of experts in eyewitness testimony (e.g., Cutler, Penrod & Dexter, 1989; Hosch, Beck & McIntyre, 1980). For instance, Hosch et al. (1980) found that participants given general information by an expert witness as to the *potential* unreliability of eyewitnesses "lowered the importance of the eyewitness testimony" (p. 294), relative to other evidence. Although verdicts and jurors' opinions of the *credibility* of eyewitnesses were unaffected, the expert testimony caused the participants to scrutinize and discuss all evidence for longer. The authors argued that expert testimony was not a specific focus of attention during deliberations, but instead helped the participants to place appropriate weight on competing evidence. Cutler et al. (1989) also found that expert testimony increased the sensitivity of jurors to factors involved in eyewitness evidence without affecting belief in the accuracy of identifications.

Until similar research has been conducted to investigate the credence given to expert testimony on jury decision making in cases involving identification from CCTV, it is not possible to assess its influence. However, if results are comparable, it is probable that jurors will place a greater, weight on their potentially mistaken own ability to view any resemblance to the defendant, than the testimony by experts in photographic analysis.

**10. 2: Photographic comparison issues**

There are no agreed methods or standards for comparing two images of the same face, although as discussed in Chapter 2 there is extensive literature describing computer algorithms for automatic face identification (e.g., Brunelli & Poggio, 1993; Heisele et al., 2003; Wiskott et al., 1997). However, Sinha (1998) argues that computer engineers have concentrated on the pattern recognition component of facial identification and not on the specific requirements of forensic image analysts required to match two images. Therefore, until standards are more accurate, it is likely that a human visual analyst will still be required to perform a match between CCTV images and photographs of the defendant when giving evidence in court.

There are three general *forensic* approaches to the problem, often described as *facial mapping*. The first, *morphological classification analysis* is a method by which facial features are defined and classified based on shape and size to provide an indication of whether these properties are similar in both images. Secondly, with *photo-anthropometric analysis*, facial landmarks in both images are identified and the distances and angles between them are calculated and compared. Finally, with *photographic video superimposition*, practitioners superimpose one image over the other on a screen and perform a series of visual tests for the determination of differences or similarities. These different approaches are not exclusive. Practitioners often combine all three, dependent on availability and quality of the evidence. Because of this, Iscan (1993) argues that the photographic analyst is required to 'reinvent' the specific methodology used in every case.

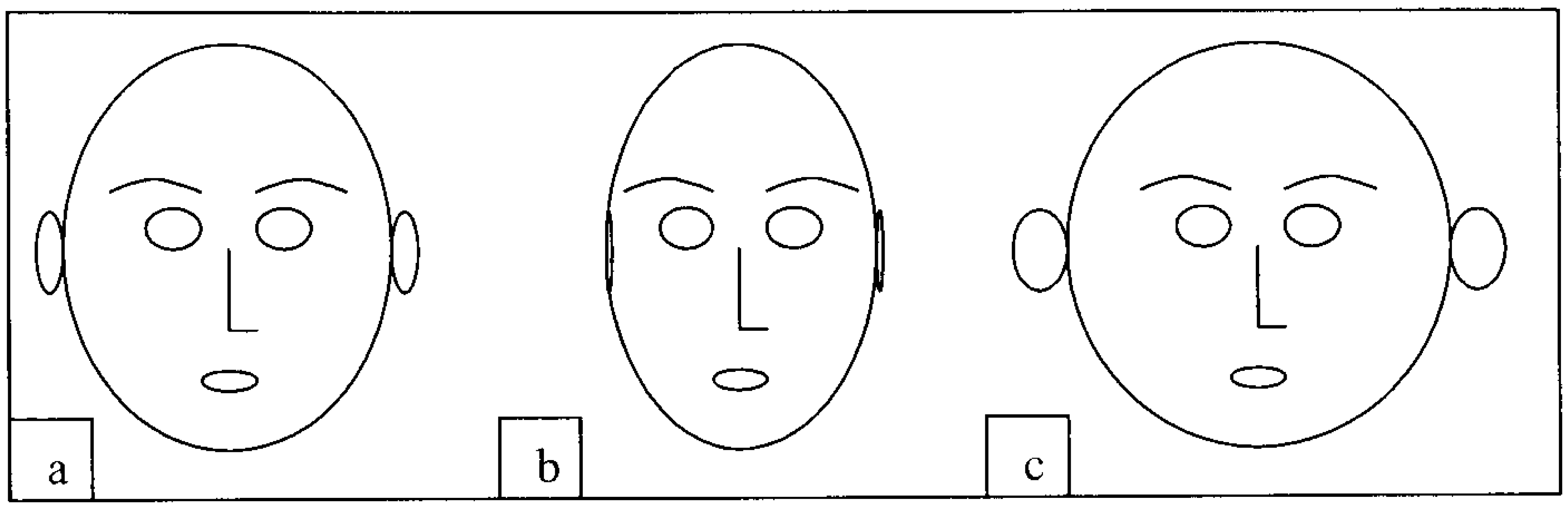



*Figure 10.1: Exaggerated schematic drawings illustrating the perceived output from different lens types while keeping head-height constant (a) standard lens from 1m; (b) extreme wide-angle lens from close-up; (c) telephoto lens from 3 – 4 m.*

One of the primary issues when faced with a photographic comparison analysis is that a two-dimensional photograph is only a representation of the underlying three-dimensional properties of a face, and depending on camera angle, features will vary in prominence. Therefore ACPO (2003) recommend that as close as possible, images provided to analysts should be taken from the same viewpoint. However, discrepancies in source equipment, brightness range, colour capture and reproduction differences can cause dispute. Indeed, Harper and Latto (2001) demonstrated that distortion effects can occur from the use of different lenses, adversely affecting all three image comparison methods. This is illustrated using exaggerated schematic images in Figure 10.1. The 'normal' face on the left (Figure 10.1a) would be as captured by a standard lens from approximately 1 metre. Close-up images from a wide-angled lens will induce concavity to an image, (Figure 10.1b) and a telephoto lens can also induce size misperceptions. From close-up, this will have the effect of widening the external features of faces (Figure 10.1c). In all these cases, if perceived head height was constant, horizontal and vertical co-planar distances would be distorted.

A further issue is that CCTV images are often of poor-quality. Bramble, Compton & Klasen (2001) note that specific software filters can refine visual data to clarify details and when correctly applied, edge detail in particular can be enhanced, although there will always be a limit. It is also possible to apply *frame averaging* techniques to multiple consecutive frames to produce one higher-quality image. This can clarify static shadowed details by equalizing the illumination of images across frames. *Frame fusion* or *frame stabilization* software also resolves issues of motion and focus blur in multiple frames, so that a stable image of a target object or person, better than that from any single frame, can be acquired. However, excessive manipulations to photographic images may be challenged as to their probity in court.

Photographic comparison analyses are mainly performed by eye, although optical devices such as a stereoscope can be used to enhance images. This creates an artificial 3-D representation and is normally applied across two adjacent still frames, as the slight change in facial position gives an impression of depth. Proponents of this technique claim that extensive training is required, as each feature on the face, its size, tone and shadow needs to be expertly defined. Furthermore, the more experienced the practitioner, the greater the perceived enrichment of the image. However, criticism has been directed at the methodology for being subjective in nature and for the inability to demonstrate in a courtroom, techniques undertaken in a laboratory (Oxlee, ND).

### 10. 3. Morphological classification analysis

One photo comparison technique is the morphological classification of facial structures. At a fundamental level this involves the categorisation of faces into population phenotypes (e.g. Mongoloid, Caucasoid etc.), and local or regional sub-divisions originally claimed by 19th century anthropologists to differentiate ethnicity based on physiognomic varieties. Until recently, the most common application of this technique was probably for the identification of human remains, with varying degrees of success. For photographic analyst and forensic purposes, feature-by-feature classification is performed, often using the same grading systems as anthropologists.

Vanezis, Lu, Cockburn et al. (1996) conducted an examination of the reliability of one method of morphological classification analysis. For this, seven raters graded high-quality photographs of five different views of 50 male faces aged from 18 – 60 years. This required the sub-classification of 39 feature categories into 87 different descriptors. For instance, there were three basic categories used to describe nose shape; - nose tip shape, nostril visibility and nasal alae. For nose tip shape there were seven descriptors; - undecided, pointed, bilobed, hooked, rounded, pronounced and asymmetrical, whereas for nostril visibility there were five descriptors and for nasal alae there were six descriptors respectively. Following data screening, fourteen categories were found to possess no discriminatory power or were associated with inter-assessor disagreement and were removed from further investigation. The authors suggest that the remaining categories might be appropriate for use in cases of disputed identification. However, no statistical analyses to individuate different faces were conducted, although, any such examination would have required nominal level inferential analyses, meaning that statistical power would have been limited. Furthermore, the database was fairly small; containing a heterogeneous sample in terms of age range and it would be unlikely that many of the faces would be the subject of identification disputes.

Vanezis et al. (1996) suggest that morphological classification is most appropriate when images are of low resolution or are taken from dissimilar angles precluding the use of other facial comparison techniques. However, they note that the technique is less effective with 'average-type' people, as they tend to be classified into the same sub-categories. Furthermore, Iscan (1993) observes that features that successfully discriminate one ethnic population from a specific geographical region may not adequately individuate those from another. Moreover, no large-scale databases have been compiled to provide an indication of the likelihood of two or more individuals possessing the same morphological characteristics. Indeed, at least one conviction has been overturned by the later discrediting of an expert witness whose testimony was based on this methodology, due to the lack of the "probability of occurrence or combinations of occurrence of particular facial characteristics" (*R v Gray*, 2003).

### 10. 4. Photo-anthropometric analysis

The second technique, photo-anthropometric analysis can be defined as the analysis of anatomical "landmarks, dimensions and angles to quantify facial characteristics from a photograph" (Iscan, 1993; p. 59). A direct comparison of the exact distances between landmarks on the face in two photographs is not normally conducted as object size can be hard to establish. This will vary as a function of the distance from the camera to the object and the specific lens used. Indeed, it is surprisingly difficult to even provide a good estimation of full-body height from a photograph (Bramble et al., 2001). Therefore, proportional analyses of the relationship between features in one image are compared with those in a second. The approach has similarities to that used in fingerprinting analysis, in that it is assumed that faces have individuating characteristics, and that if a defined number are matched in two images, identity is probable. However, unlike with fingerprints, there is no accepted methodology for the number of responding characteristics necessary for declaring a match and no accepted methodology for expressing the degree of confidence when a match is determined.

Furthermore, Mardia et al. (1996) purport that it would be a mistake to draw close analogies between facial mapping and fingerprint analysis as in the latter the topology of shape structures are well-defined, even if a print is distorted. In contrast, there are no similar highly-defined connections within a face and expression changes will alter the relative position and dimensions of the majority of facial structures. Indeed, they note that the main problem is that without knowing how common a specific set of dimensions are within the population, "the uniqueness of the individual is not known" (p. 4), meaning that it is difficult to achieve a safe conviction in the courts 'beyond reasonable doubt'. It is easier to eliminate a suspect than to make a positive identification, as if one reliable difference is found that suspect can be excluded.

Indeed, following a number of court judgements (e.g., *R v Gray*, 2003) there have been calls for a national database of facial measurements so that specific proportions in the population can be calculated to give an indication of the likelihood of two different individuals having similar face structures. However, Iscan (1993) argues that a particular set of landmarks cannot be standardised as it depends on those available within any photographic evidence. This will often be determined by the specific facial viewpoint captured by a camera. Furthermore, recent research has demonstrated that even fingerprint analysts are susceptible to contextual information (Dror, Charlton & Peron, 2006). If misleadingly told that it was unlikely that two fingerprints were of the same person, experts unsuspectingly provided contradictory 'no match' judgements when previously positive identifications had been made when no such contextual information was given. It is possible that all facial analytical methods would be vulnerable to cognitive bias effects of this type.

Some research using photo-anthropometrical measurements has been published (e.g., A.M. Burton, Bruce & Dench, 1993; Catterick, 1992; Laughery et al., 1981; Mardia et al., 1996). For instance, Mardia et al. (1996) conducted a series of analyses using a database of 358 young white male faces, captured in full-face and profile views taken in a controlled environment. Twenty landmark distance measurements were collected, as were the angles between these landmarks to conduct shape analysis. The authors found that there were high correlations between all measurements indicating limited ability to distinguish between different faces on the basis of these 11 full-face and 9 profile distances. However, profile and full-face viewpoint analyses were conducted separately and if data were combined, a more robust method of distinguishing between faces may have emerged. Nevertheless, this research illustrates the difficulties involved in applying the technique with even extremely high-quality viewpoint-standardised images. Most CCTV facial images will be of far lower quality, with variations in camera viewpoint, enhancing the potential for error.

**10. 5. Photographic video superimposition**

The final technique used for image comparison is photographic video superimposition. This requires specialist equipment and involves the superimposing of two projected images over one another in order to apply various fading mechanisms to "make one face disappear into another, with the second image eventually replacing the first" (Iscan, 1993; p. 63). Specific techniques include visual flicker, and vertical, horizontal or diagonal wiping so that a line erasing part of one image reveals part of the second, allowing similarities, or more saliently discrepancies between the two images to become apparent. Indeed, Sinha (1996) argues that used with image enhancing computer software, superimposition can reveal symmetry or a lack of symmetry across images.

However, superimposition is extremely susceptible to changes in facial viewpoint and a number of 'camera tricks' can make facial images appear identical. For instance, the slower the fade the more likely an 'illusion' of a perfect match will be achieved, so that when presented in court it can provide highly persuasive evidence. Indeed, large alterations can readily be introduced to visual scenes without any perceptual awareness (Simons & Ambinder, 2005: Simons & Lewin, 1997). Called *change blindness*, this psychological phenomenon has been demonstrated using many different types of stimuli and even the substitution of one human with another in film footage can easily be missed, suggesting an inability to retain information from movement to movement. Therefore, evidence presented in court based on visual superimposition should be treated with great caution.

However, the recent development of photographic equipment that can acquire three-dimensional (3-D) images has led to suggestions that these could be used in forensic investigations in conjunction with both superimposition and photo-anthropometric techniques. For instance, Yoshino, Matsuda, Kubota et al. (2000) using a 3-D physiognomic range finder, demonstrated that a two-dimensional extract can be accurately superimposed over a target image captured from a conventional camera. To ensure that viewpoint was equivalent, seven anthropometrical locations were marked on both images. Software then automatically adjusted the 3-D range finder image to match that of the 2-D camera image by calculating the average perpendicular distance between each point. Yoshino et al. describe a method of applying simple anthropometric analyses by calculating the reciprocal point-to-point differences between both images. In a later demonstration, Yoshino, Noguchi, Atsuchi et al. (2002) applied this technique to a database of 100 faces, in which novel faces were entered as probes. The authors claimed a 100% identification rate as the measured differences in two different images of the same person were always less than those of two different people. However, the age range of faces was 24 – 46 years and no details were given of their perceived similarity, making it unclear whether any would be mistaken for one another by human observers.

Yoshino et al. (2002) suggest that ideally 3-D images could be acquired of suspects in a similar manner to normal police mug-shot photographs. The technique could then be routinely applied when security footage of an incident is obtained, by comparing the images to a 3-D facial database. However, most 3-D technologies suffer from image distortion effects in part due to lighting anomalies but also because of slight inadvertent movements, as image acquisition can take several seconds (Schofield & Goodwin, 2004), limiting its use with uncooperative suspects. However, recently the Home Secretary, Charles Clarke was reported to have inspected a system able to capture 3-D images in less than 50 ms (Reed, 2006), and further technological advances may reduce some associated image deformations.

Evidence from a video superimposition expert witness was first admitted in court in the UK in the early 1990's, with the technique's status confirmed on appeal (*R v Clarke*, 1995). However, it has been argued by the defence, and repeated by the judge when summing up in one particular case that it:

> "Is really just a subjective assessment, it is not scientific, he is just a man with a magnifying glass. There are no measurements or calculations or anything of that kind", (*R v Kerrigan*, 1998).

Indeed, the methodology itself has been criticised as analysts claim to be able to 'see' details in visual images that are invisible to the untrained eye because of their 'experience and equipment' (e.g., *R v Gray*, 2003). Nevertheless, this appeal was dismissed despite other experts arguing that the evidence was unreliable and measurements had not been compared against a facial database. The court ruled that the methodology used had allowed the witness to provide opinion evidence based on his ability to determine correspondence between images in his lab. This allowed him to guide the jury as to the detail contained within the images in a similar manner to the ruling applied in *R v Clare and Peach* (1995) in which a police officer had provided ad-hoc expert testimony based on his own visual inspection of CCTV footage.

More recently, the same court has ruled that knowing the likelihood of shared facial characteristics is not necessary when providing this type of evidence (*R v Gardner*, 2004). In their ruling, judges noted that if a technique could be shown to aid the jury, an experienced practitioner using specialist equipment may present their *subjective* opinion of identity in court, based on their personal observations. However, professionally presented prosecution evidence of this type can be extremely convincing, appearing to be scientifically based. In this particular case, defence experts argued that it was not possible to substantiate the specific claims made by the prosecution, in particular because the quality of footage was poor. However, the Court of Appeal ruled that the judge had correctly cautioned the jury that care should be taken when evaluating the prosecution expert's testimony and the conviction was allowed to stand.

### 10. 6. *DigitalFace* photo-anthropometrical software

There are a number of unresolved issues concerning the reliability of photo-anthropometrical analysis. However, it is the only photo comparison technique to allow highly detailed measurements of facial structure and is therefore likely to remain the focus of research within forensic science. If used in conjunction with 3-D image capture and superimposition technology, and the development of suitable databases of facial measurements, it may provide influential evidence in court. Indeed, unlike other methods, error levels can also be assessed and parametric analyses can be conducted.

Therefore, to conduct the final series of studies reported in this thesis, a custom software-assisted facial landmark identification system, *DigitalFace* was specifically designed to carry out exploratory research into the forensic reliability of photo-anthropometric analysis. The writer of this thesis provided the basic specifications for this system. However, Rob Davis from the information technology department of Goldsmiths College, University of London designed the software.

The system requires an operator to locate and denote (*digitize*) 38 specified landmark sites in full-face (*anterior*) view; and 14 in *profile* view on facial images displayed on a computer monitor. Once identified, the system produces a database of 37 linear and 25 angular measurements between those sites. The landmarks chosen replicated and extended those used in previous anthropometrical (Catterick, 1992; Mardia et al., 1996; Laughery, Rhodes & Batten, 1981) and psychological studies (e.g., A.M. Burton et al., 1993; Rhodes, 1988). Rhodes (1988) in particular had specifically selected orthogonal measurements designed to delineate the size and shape of facial features, as well as the spatial relationship and location of those features within each face. However, some of the features measured by *DigitalFace* lack permanency, such as eyebrow structure or hairline, and it may not be appropriate to include them in all forensic investigations. These are referred to as *transient* measurements, as opposed to the remaining *permanent* measurements throughout this chapter. *DigitalFace* will operate most effectively if facial images are captured directly from the front or side. However, this would not preclude its use with images from other angles as long as all those being compared were from the same viewpoint.

There were two main aims of this research. The first was to test whether the system could successfully discriminate between two images of two different individuals. The second was to assess whether it could identify that two different images depicted the same person, and to provide a measure of confidence in that decision. A series of analyses were conducted as though each was a separate photo-anthropometrical forensic investigation. The methods used by witnesses giving expert evidence of this nature in court have been criticised if the expert has stated that facial dimensions match in both images, without providing any indication of how many other people in the population could possess a similar structure (e.g. *R v Gray*, 2003). Therefore each analysis involved the establishment of a minimum acceptance criterion for identity determination, tested against the already digitized database of 100 frontal (*anterior*) and 70 profile homogeneous facial images described in Chapter 3 (anterior images are depicted in Appendix E).

The first series of analyses examined whether novel photographs (*probes*) of faces already stored within the database (*targets*) would be identified as being of the same person. To establish a minimum criterion for a match if multiple images from different views had been obtained, separate examinations were conducted with anterior and profile databases and to test system flexibility with a limited set of parameters, disguised images were also included. Furthermore, analyses were conducted with and without transient features such as the structure of the eyebrows and the hairline. It would be unlikely that in a real photo-anthropometrical investigation, comparison images would be obtained on the same day and from a posed viewpoint. Therefore, the undisguised probe images were profile and anterior views taken of the eight actors at the time of the identification sessions for Experiment 6.1, approximately three weeks after the database photographs were taken (see Figure 6.1). The same camera was used; however, there were hairstyle variations and they were not so strictly posed, meaning that viewpoint was not exactly matched. Nevertheless, the quality of the images was higher than would normally be captured by CCTV, allowing a test of the system in optimal conditions.

With the *DigitalFace* system, it would be possible to implement a simple decision rule by examining whether the correlation between probes and target was higher than that between any two faces on the database. However, the *lowest* correlation coefficient between two different anterior view faces in the database on all dimensions was above $r$ = .98. Probably this was partly due to the homogenous inclusion criteria for the database, but this also demonstrates the high degree of similarity in the structure of human faces. Therefore, hierarchical cluster analyses using the entire set of database faces were conducted, and decisions were based on the distance in Euclidean space as defined by generated coefficients within proximity matrices.

Absolute distances can rarely be measured in a photograph, without knowing the exact camera distance and lens focal length (Bramble et al., 2001; Iscan, 1993). Therefore, in court, photo-anthropometrical analysis will involve the presentation of landmark measurements defined as the standardised ratio of a *referent* distance between two specific landmarks. In anterior view, this will often be the distance between the top of the head and the chin for vertical dimensions and the distance between the outside of the ears for horizontal dimensions. However, for forensic purposes the selection of a referent will need to be flexible and will depend on available landmarks. Indeed, in many of the images in the database, the top of the head, due to hairstyle, was hard to define. Therefore, different measures were utilised, in particular for when images were depicted in disguise.

In addition, many of the measures will be highly related to one another, such as the distance from the top of the upper lip to the chin and the distance from the lower lip to the chin. Therefore, to avoid redundancy, all distances were converted to new independent *vectors* for all analyses. In the example above, one vector would be the distance between the upper and lower lips, the second vector the distance from the lower lip to the chin.

For a positive identification decision, there were two criteria. The first was that the vectors of two images of the *same* person (probe and target) should be closer in terms of Euclidean space than the distance from the probe to any *distracter* within the database. The second, more rigorous criterion was that the distance in Euclidean space between these two images of the *same* person should also be less than that between all other pairs of images of two *different* people within this database.

A further examination tested the reliability and consistency of landmark identification using the *DigitalFace* system. For this, a single probe face was digitized 80 times, with various settings of image enlargement on the monitor. This was to examine the potential error associated with images of the same person that may be captured from different distances. A similar hierarchical cluster analysis was conducted with the data from these varied settings of 'zoom' entered as further probes. In addition, the values equal to the 95% confidence limits, 1 SD and 2 SD above the mean on each of the distance and angular vectors were added as additional probes within this analysis. In this manner, the error associated with landmark identification was transformed into a measure of error within Euclidean space. In all of these cases, it was expected that the Euclidean distance between the data derived from the *same* image would be less than between two *different* photographs of the *same* person.

Finally, some of the different faces stored in this database had been incorrectly identified as the same person by human participants in the matching experiments reported within this thesis. These experiments replicated scenarios that might occur in real life, and if these particular actors had actually been the subject of a criminal investigation, expert witnesses may have been requested in order to ascertain identity. This database therefore enables the versatility of *DigitalFace* to be explored in a forensically-relevant investigation. Commentary is included with the results of the analyses described above, if, on the basis of the specific parameters incorporated in each, these particular actors might also be mistakenly matched using the *DigitalFace* procedures. In addition, a final analysis applied the technique to pairs of faces that in each separate experiment conducted within this thesis had been the subject of the *highest* error rates, to examine the effectiveness of *DigitalFace* to disambiguate perceptually similar individuals.

### 10. 6. 1. Method

#### 10. 6. 1. 1. Facial images

The database of 100 anterior and 70 profile black-and-white facial images were those described in Sections 3.3 and 3.7. All frontal view images from this database are presented in Appendix E. All images were of white Caucasian males aged between 18 and 25 meeting the experimental inclusion criteria described in Chapter 3. The undisguised probe images were separate photographs of the eight actors recruited for the live actor matching studies reported in Chapter 6. The database target and probe images of these actors are depicted in Appendix D.

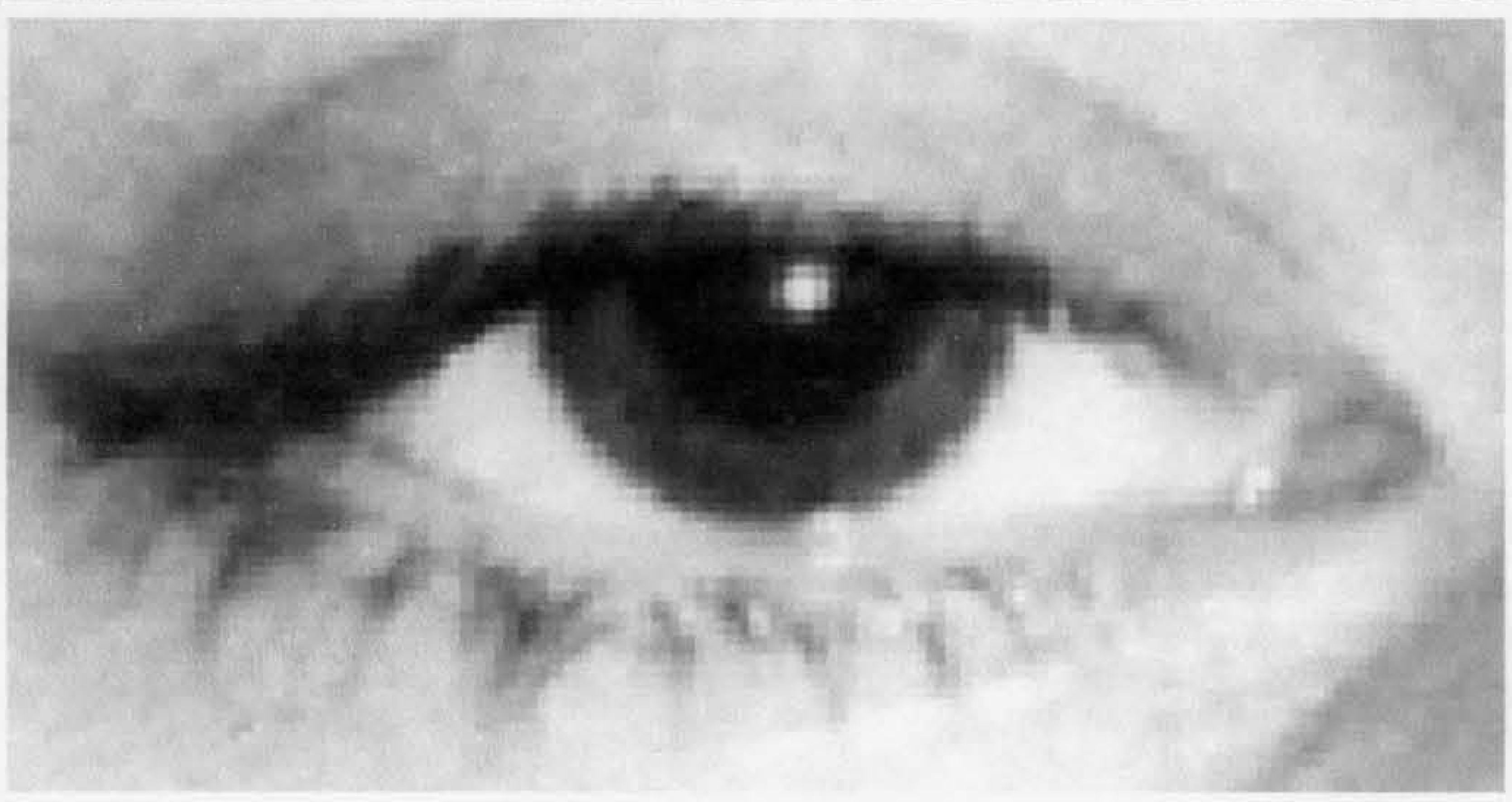
The second set of probe images, to examine the effect of disguise were of the same eight actors from anterior and profile views wearing a hat, taken on the same day as the undisguised images stored in the database (Appendix D).

To measure the error associated with facial landmark identification using *DigitalFace*, the anterior image of Actor 44 who took part in Experiment 7.1 was employed (Appendix E).

#### 10. 6. 1. 2. Procedure and landmark database

Each database and probe facial image was 'digitized' eight times by the same experimenter using the *DigitalFace* system on a 16" monitor screen. Within the system, if required, facial images can be enlarged and reduced in size at any time, although maximum enlargement is dependent on resolution. In forensic scenarios it may be of use to zoom in or out to correctly identify the location of individual features. This is because with extreme close-up images, some landmarks are harder to define than when using longer distance views and vice-versa. In this case, digitization was conducted four times with the images on full zoom (zoom x 3) and four times on mid-zoom (zoom x 2). In the former, with images from the database, the average apparent distance on the screen between the pupils of the eyes was 15 cm. With the latter, it was approximately 10 cm, both larger than in real life.

For the landmark identification error and confidence level analyses, the same anterior image (Actor 44) was digitized 80 times by the same experimenter on different days over a period of 2 weeks. Each consecutive digitization procedure was conducted with a different setting of zoom on the monitor screen (no zoom: apparent distance between eyes = 4.2 cm; zoom x 1: 6.3 cm; zoom x 2: 9.5 cm; zoom x 3: 13.5 cm). Figure 10.2 demonstrates the exact size of the left eye as displayed on the screen in the highest and lowest zoom settings.



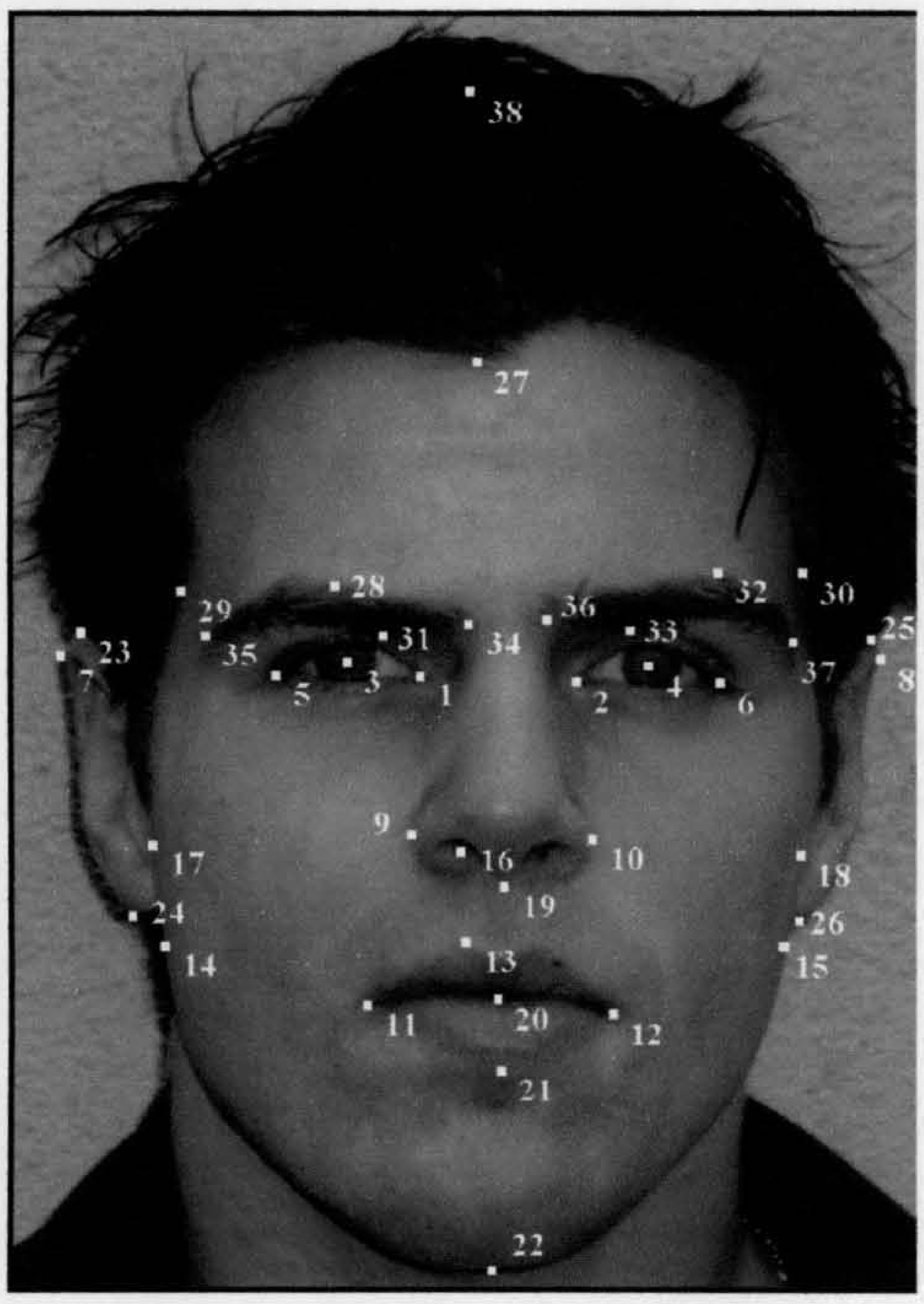
*a)*

*b)*

*Figure 10.2: Perceived size of eye at a) zoom x 3 setting and b) no zoom*

Throughout the process, *DigitalFace* provides in the toolbar a visual reminder to the operator of each feature to be identified using its 'common' name in the order listed in Figures 10.3 for anterior views and 10.4 for profile views. On the screen, cross-hairs are visible and with a computer mouse these can be moved until the target landmark is located. Clicking on the left and then the right mouse saves that location. Some landmarks can only be located with the use of a horizontal or vertical projected line. For instance, measurement of the width of the face at various vertical distances is performed by the initial selection of an internal landmark (e.g. top upper lip; *superior labiale*). From this, a line parallel with the inner eyes (left and right *endocanthian*) is projected onto the screen for simple identification of the edge of the face at this vertical height. The entire procedure is simple, requiring little training as instructions are always available. Once fully trained, each image takes an operator less than 5 minutes to digitize.

Following final landmark identification, *DigitalFace* automatically exports pixel distances and angular measurements into a previously-defined database. This database contains 25 distance and 14 angular measurements in anterior view and 12 distance and 11 angular measurements in profile view. A full list of these measurements as well as the figures plotting locations can be found in Appendix C.1.1. Data screening was conducted on this database and any outlier more than 2 SD from the mean for that face was removed. These denote obvious landmark identification errors. The median value for each measurement from the remaining data was retained for all further analyses.

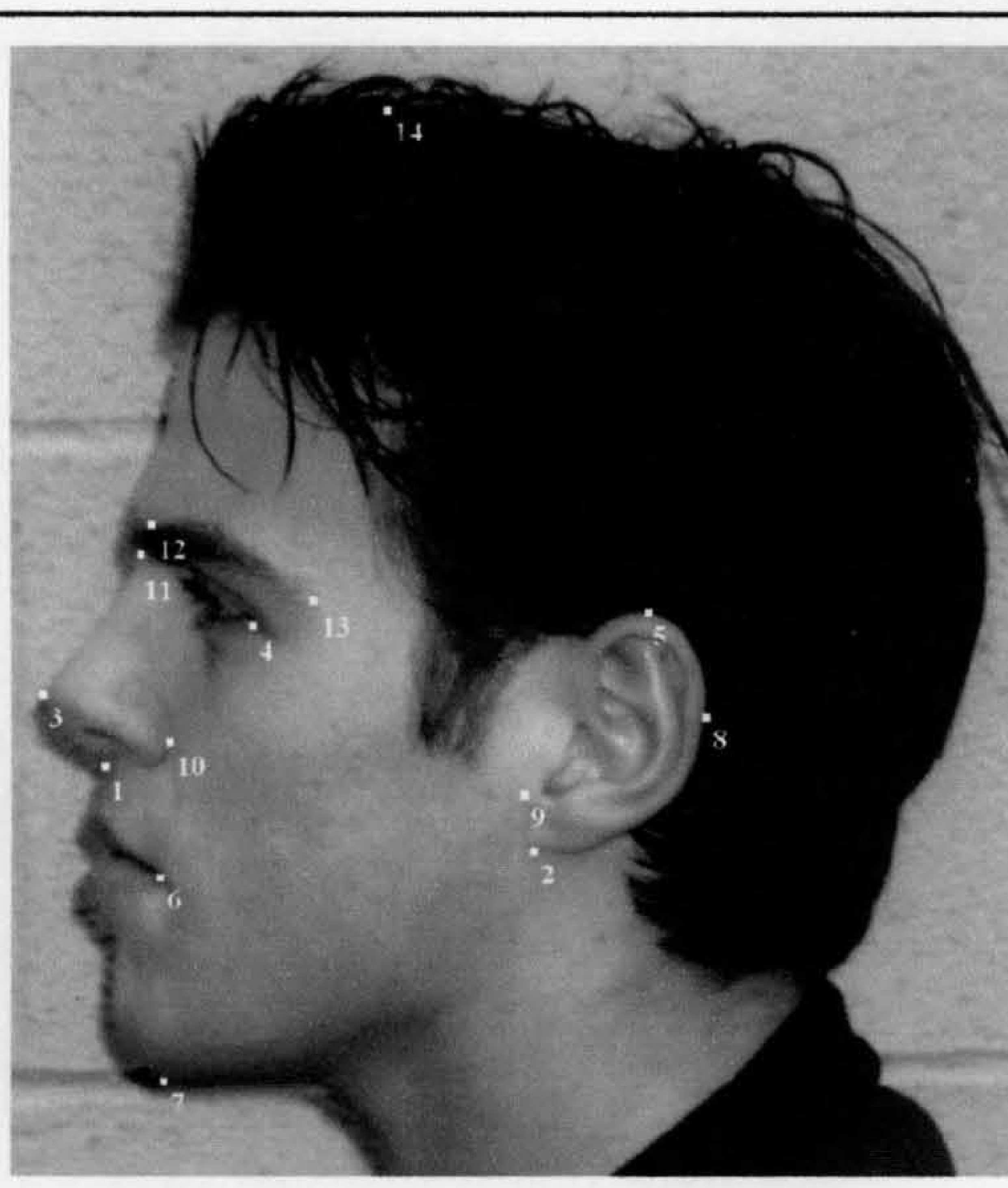


**1.** Right endocanthian (inner right eye); **2.** Left endocanthian (inner left eye); **3.** Right pupil (centre point of right pupil); **4.** Left pupil (centre point of left pupil); **5.** Right exocanthian (outer right eye); **6.** Left exocanthian (outer left eye); **7.** Right postaurale (right posterior of ear); **8.** Left postaurale (left posterior of ear); **9.** Right alare (most lateral point of right nasal area); **10.** Left alare (most lateral point of left nasal area); **11.** Right cheilion (right outer mouth); **12.** Left cheilion (left outer mouth); **13.** Left superior labiale (SL, top of upper left lip); **14.** Right lateral point of face at SL line; **15.** Left lateral point of face at SL line; **16.** Left supra subalare (SS, centre of left nostrum); **17.** Right lateral point of face at SS line; **18.** Left lateral point of face at SS line; **19.** Subnasale (bottom of nose); **20.** Stomion (centre of mouth); **21.** Inferior labiale (bottom of lower lip); **22.** Gnathion (bottom of chin); **23.** Right superaurale (top of right ear); **24.** Right subaurale (bottom of right ear); **25.** Left superaurale (top of left ear); **26.** Left subaurale (bottom of left ear); **27.** Trichion (hair line at midpoint of forehead); **28.** Right superciliare (SC, top of left eyebrow); **29.** Right lateral face on SC line at hair contact; **30.** Left lateral face on SC line at hair contact; **31.** Right orbitale superious (bottom of right eyebrow); **32.** Left superciliare (top of left eyebrow); **33.** Left orbitale superious (bottom of left eyebrow); **34.** Right inner eyebrow; **35.** Right frontotemporale (outer right eyebrow); **36.** Left inner eyebrow; **37.** Left frontotemporale (outer left eyebrow); **38.** Vertex (highest point on head)

*Figure 10.3: Locations of anterior facial landmarks: Anatomical definitions and common names given in instructions to DigitalFace. Note: right and left locations are from the perspective of the viewer*

It would be acceptable to perform a photo-anthropometrical analysis with these measures, if conducted on a limited set of images, for instance, for a court case. However, the main aim of this exploratory research was to examine their use against a database of faces to assess system reliability. Therefore, image normalization was conducted by expressing landmark measurements as a ratio of the *referent* distance between two specific landmarks.

For the undisguised face analyses, in anterior view the final converted vertical proportions were ratios of the referent distance between a line linking the inner eyes (*endocanthii*) and the chin (*gnathion*), the horizontal ratios were derived from parallel measurements to the distance between the ears (right and left *postaurale*). In profile view, vertical ratios were derived from parallel distances between the chin (*gnathion*) and the upper ear (*superaurale*). Profile horizontal ratios were derived from measurements parallel to a line linking the tip of the nose (*pronasale*) and the rear of the ear (*postaurale*).



**1.** Subnasale (nose base); **2.** Right subaurale (right ear base); **3.** Pronasale (most protruded part of nose); **4.** Right exocanthian (right outer canthus of eye); **5.** Right superaurale (top of right ear); **6.** Right cheilion (right corner of mouth); **7.** Gnathion (chin); **8.** Right postaurale (rear of ear); **9.** Right otobasion infrious (front of ear, point of attachment of ear lobe to cheek); **10.** Right alar curvature (most lateral point of the curved part of the alar of the nose); **11.** Sellion (deepest landmark at the top of the nose); **12.** Glabella (prominent midpoint of eyebrows); **13.** Right rear frontotemporale (right outer eyebrow); **14.** Vertex (highest point on head)

*Figure 10.4: Locations of profile facial landmarks: Anatomical definitions and common names given in instructions to DigitalFace*

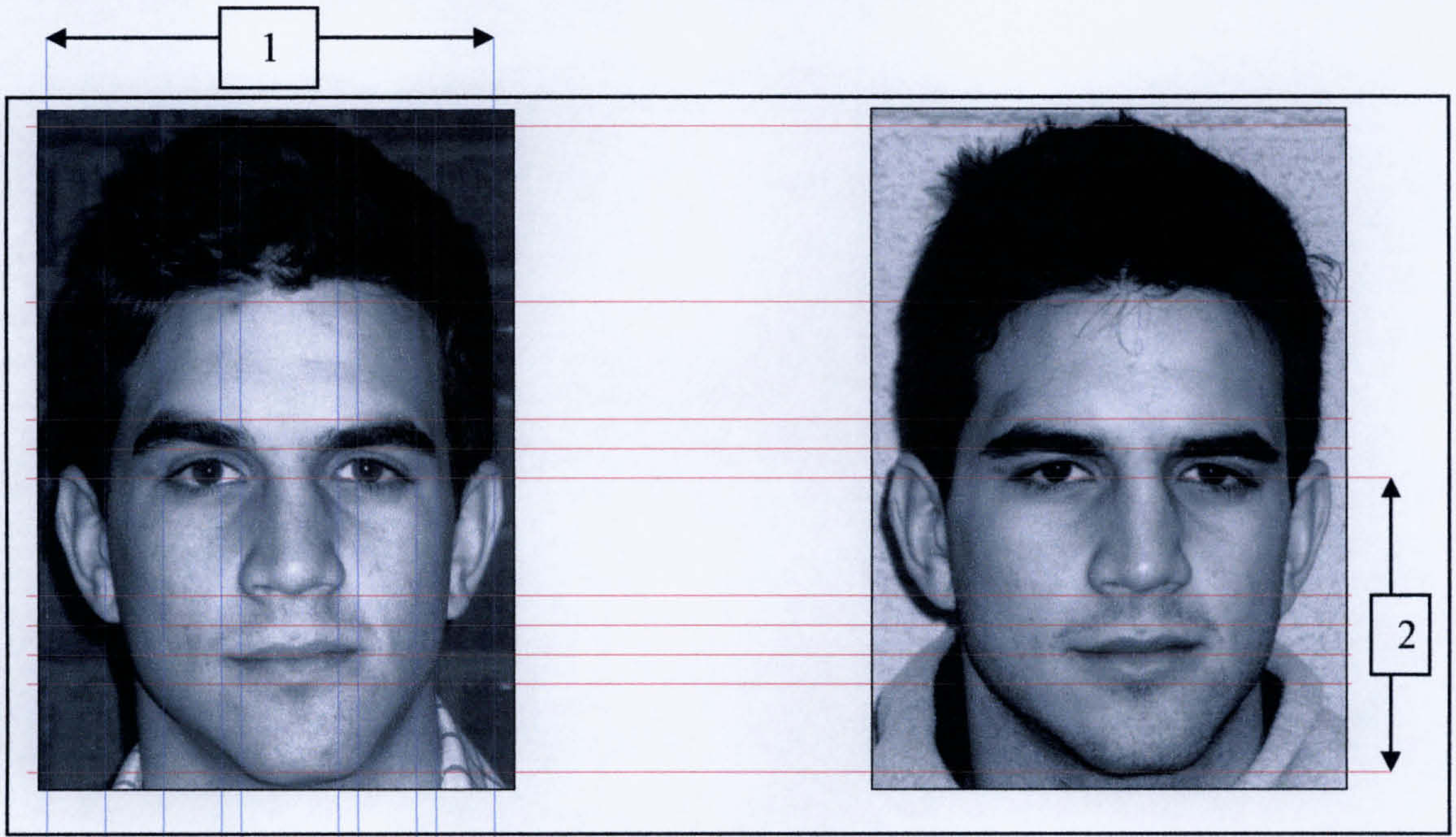
Alternative referents were required with the disguised faces. In anterior view the width of the face at the height of the left nostrum (*supra subalare*) was used for all horizontal measurements (the vertical referent distance was as above). In profile view, all horizontal ratios were derived from a referent line linking the front of the nose (*pronasale*) and the rear of the nose (*alar curvature*).

However, many of these measures are highly related to one another. For instance, measurements automatically calculated by *DigitalFace* include the distance from the bottom of nose (*subnasale*) to the bottom of the lower lip (*inferior labiale*) the subnasale to the upper lip (*superior labiale*) and the subnasale to the centre of the mouth (*stomion*). Therefore, to avoid redundancy, all distances were converted to new independent *vectors* and standardised so that the unique variance within each could be analysed. Some of these converted vectors are presented in Figures 10.5 and 10.6 as an illustration of how the results of a photo-anthropometrical analysis could be presented in court; and demonstrating the two distance referents from which all non-disguised ratios were derived. Due to space restrictions and clarity not all distance vectors are included. Furthermore, the presented lines do not exactly align with features as this is only a demonstration. However, all final *vectors* are listed in Appendix C.2, together with descriptive statistics for the entire database.

### 10. 6. 2. Results and discussion

#### 10. 6. 2. 1. Identification of probe images

A series of hierarchical cluster analyses were conducted to examine the utility of the *DigitalFace* system to identify whether two different images of the same person were more likely to be matched, than to be incorrectly identified as being the same as another face on the database. For this, probes whose target faces were already stored on the database were entered into each analysis, together with all *distracter* faces on the database. A separate proximity matrix was generated for each analysis. A reliable test should find that probe faces had a lower squared Euclidean distance to its respective target, than to any other face within the database. This value should also be lower than for any two different faces.



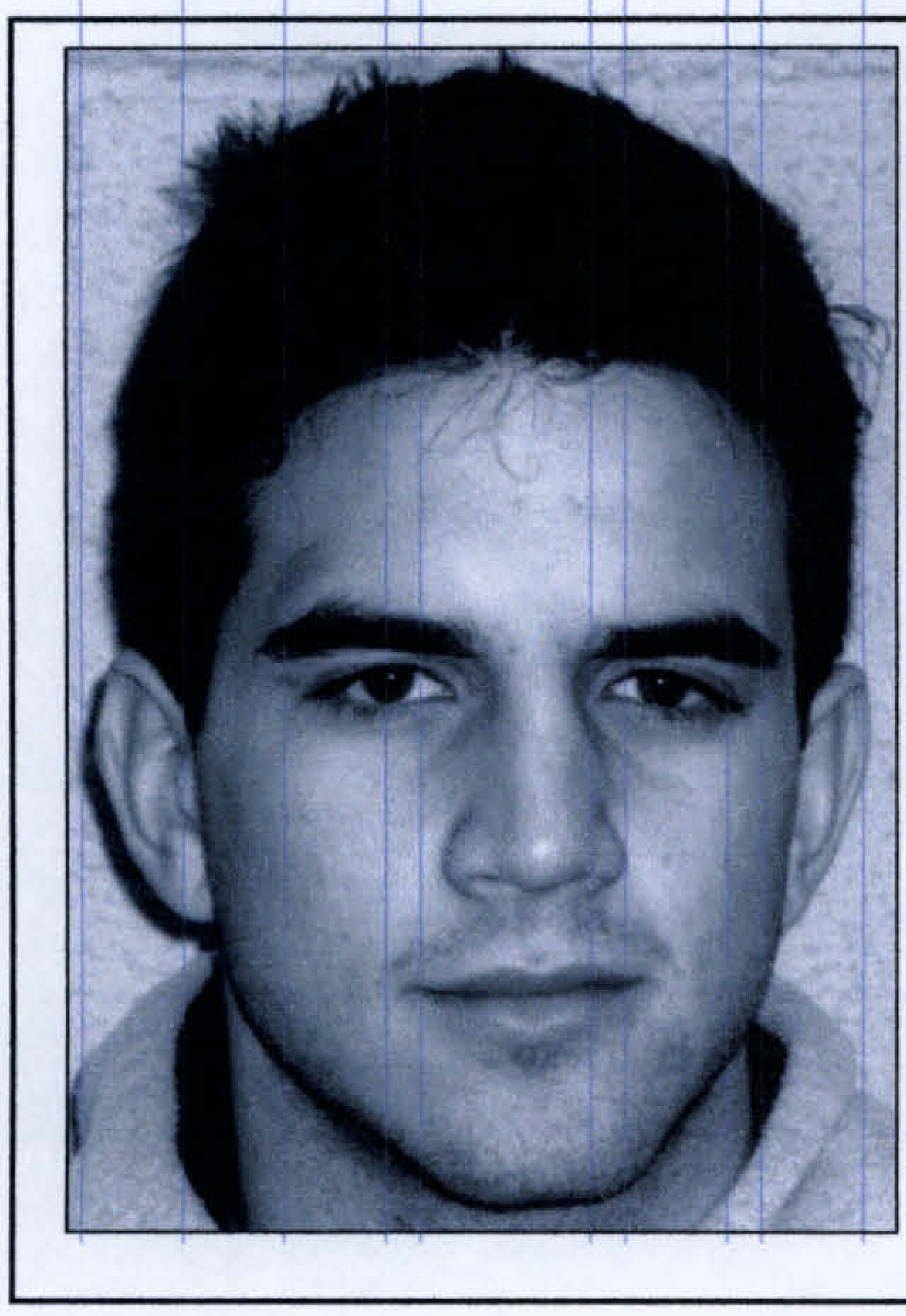



*Figure 10.5: Illustration of how the results of one of the photo-anthropometrical analyses conducted in Chapter 10 could be presented by an expert witness in court. Accompanying the figure would be a table detailing the measurements in terms of* ***1.*** *Horizontal distances, expressed as ratios of the distance between the superaurale (ears), and* ***2.*** *Vertical distances, expressed as ratios of the distance from the gnathion (chin) to the inner canthii (eyes). In this case, the same actor (A02) is depicted in the images. The top left photograph was taken at the time of the identification sessions for Experiment 6.1; the other image was obtained when the video was filmed three weeks previously.37.5% of participants made errors in the target present trial involving this actor.*

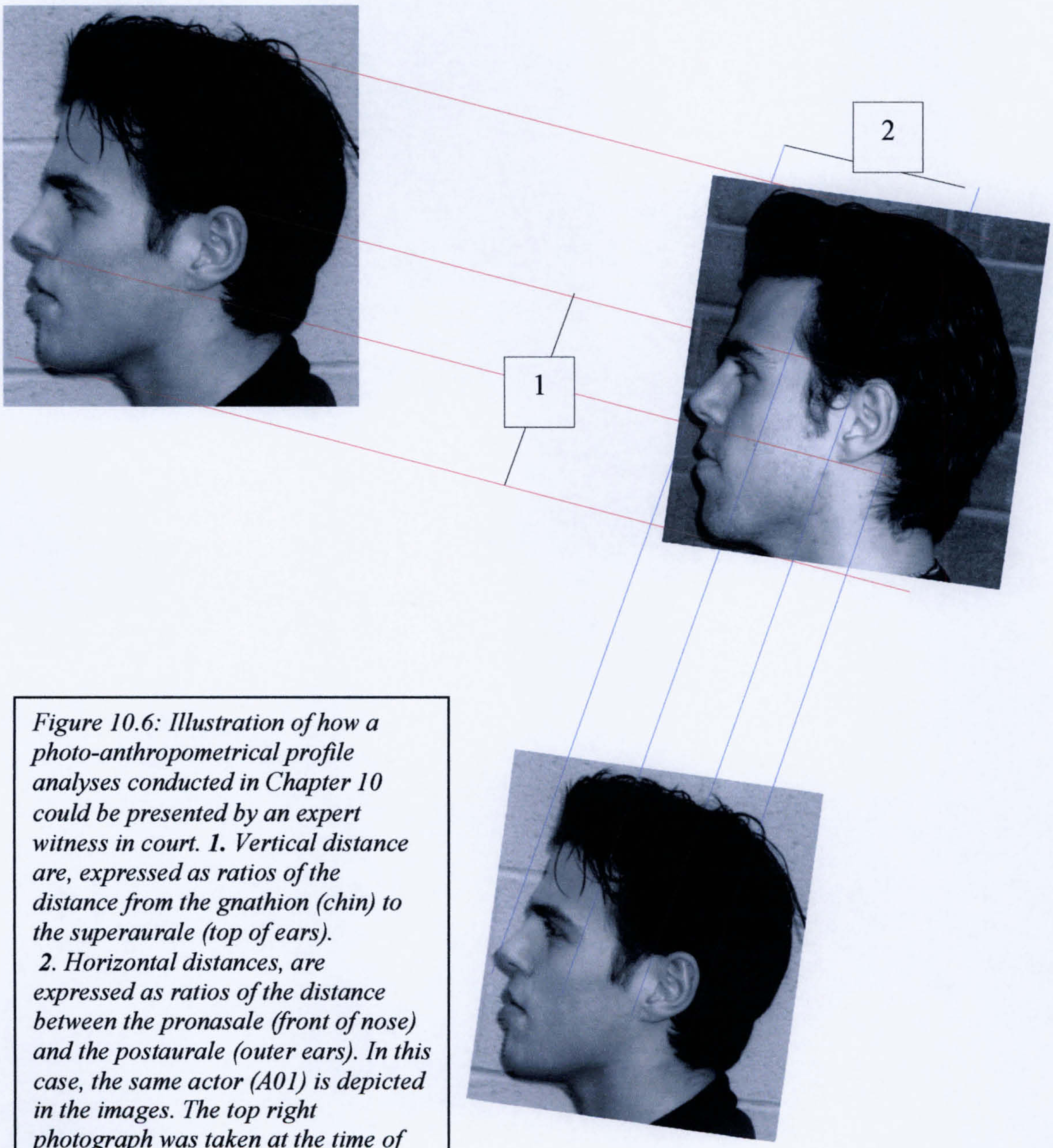



*Figure 10.6: Illustration of how a photo-anthropometrical profile analyses conducted in Chapter 10 could be presented by an expert witness in court.* ***1.*** *Vertical distance are, expressed as ratios of the distance from the gnathion (chin) to the superaurale (top of ears).* ***2.*** *Horizontal distances, are expressed as ratios of the distance between the pronasale (front of nose) and the postaurale (outer ears). In this case, the same actor (A01) is depicted in the images. The top right photograph was taken at the time of the identification sessions for Experiment 6.1; the other image was obtained when the video was filmed; three weeks previously. None of the participants made errors in the target present trial involving this actor.*

For each of the following tests, the associated table lists the squared Euclidean distance between each probe and target, as well as the lowest value between that probe and any distracter within the data set. This provides an indication of the likelihood that on the particular set of included vectors it would be mistakenly matched with another. Furthermore, the mean (with standard deviation) squared Euclidean distance for the entire database is listed, together with the maximum and minimum distances between any two *different* faces within the database. The latter provides a robust minimum criterion for all positive decisions concerning identity.

In addition, data from the subjective facial similarity matrix, described in Section 3.7 are included. This provides an indication of whether the probe face and the distracter possessing the most similar physical structures as measured by *DigitalFace* would be likely to be mistakenly matched by human observers. With 75 pilot participants, the maximum number of potential matches was 75. However, values varied from 0, meaning that none of the pilot participants matched those two faces; to 25, the highest within the database. The mean number of matches for the entire database of 4,950 cells was 0.75 (*SD* = 1.64). Twenty-six pairs (.005%) of faces were matched ten times or more, 111 (2.24%) matched 5 times or more and 3,100 (62.3%) pairs of images were matched by no participants.

**10. 6. 2. 2. *DigitalFace* landmark identification and associated error analysis**

The first hierarchical cluster analysis examined the variation associated with landmark identification using the same facial image (Actor 44), using all the database of anterior distance and angular vectors. The mean value for each distance and angular vector from the 20 digitization processes at each level of zoom was entered as a probe into the analysis. In addition, the values equal to the 95% confidence level, 1 SD, and 2 SD above the mean from all 80 digitizations were also added as additional probes. The results are presented in Table 10.1.

The squared Euclidean distance of all the different zoom settings is less than 1 unit from the mean, indicating that the identification of landmarks did not vary much as a function of the relative size of the image on the screen. Furthermore, the minimum distance between a probe and a target of the same person (5.04; P06) is more than the distance between the mean and the vectors derived from the 95% confidence levels (0.29) and those derived from a distance 1 SD from the mean (2.25). However, the squared Euclidean distance of all the vector values 2 SD from the mean is higher (8.99), indicating that if landmark identification is less robust, errors could be made if attempting to determine if two images were identical (perhaps if resolution was poor). Nevertheless, the minimum distance between two images of different people (14.85; A09; A10) was much higher, suggesting that digitizing the same impoverished image would be reliable at ensuring two different people were not depicted.

*Table 10.1: Frontal anterior probe entries analysing error associated with landmark identification (Appendix C.2.1 – C.2.4)*

| From closest target (A17) | 95% confidence level | Vectors 1 SD from mean | Vectors 2 SD from mean | Distance from overall mean value | | | |
|---|---|---|---|---|---|---|---|
| | | | | Zoom x 0 | Zoom x 1 | Zoom x 2 | Zoom x 3 |
| 38.37 | 0.29 | 2.25 | 8.99 | 0.69 | 0.56 | 0.51 | 0.87 |
| Min distance between two photos of same person | | | 5.04 A06 | Min distance between photos of two different people | | | 14.85 A09 A10 |
| Max distance between two photos of same person | | | 16.49 A03 | Max distance between photos of two different people | | | 299.65 A40 A96 |
| Mean 78.00 (SD = 37.39) | | | | | | | |

### 10. 6. 2. 3. Matching anterior undisguised probes with database targets

The second series of cluster analyses examined whether the probe undisguised frontal photographs, taken of the eight actors during their identification sessions in Experiment 6.1 would be correctly matched to their target facial photographs taken three weeks previously. Separate analyses were conducted using the anterior and profile views, with and without the inclusion of angular and transient vectors.

#### 10. 6. 2. 3. 1. Matching permanent anterior distance vectors

For the first test, only the 15 permanent anterior distance vectors were included, as these would be considered the most reliable in a photo-anthropological analysis (Table 10.2).

Using these vectors, the lowest squared Euclidean distance value between two different faces (A48 & A67) in the database, at 2.48 was lower than the distance between four probes and their respective targets. In addition, 35 further cells within the proximity matrix had distance vectors lower than 6.05, (the highest distance between two images of the same person; A01), indicating that using this limited set of vectors, these pairs of faces would be more likely to be incorrectly matched, than the actual target with their probe.

*Table 10.2: Permanent anterior distance measurements only (Appendix C.2.1 & C.2.2)*

| Squared Euclidean | Probe names | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance from probe | A01 | A02 | A03 | A04 | A08 | A09 | A10 | A33 |
| Target | 6.05 | 1.33 | 0.74 | 1.88 | 5.72 | 4.87 | 2.84 | 2.04 |
| Closest distracter | 6.18 | 3.52 | 4.89 | 6.82 | 4.94 | 3.80 | 4.84 | 12.23 |
| Distracter ID code | A88 | A42 | A57 | A89 | A03 | A10 | A85 | A23 |
| Similarity matches | 2 | 0 | 3 | 2 | 4 | 7 | 1 | 4 |
| Lowest distance between any two different faces | | 2.48 (2) | A48 A67 | Highest distance between two faces | | | 145.3 (0) | A40 A90 |
| Overall database mean (SD) | | 30.69 (18.55) | | | | | | |

In addition, the squared Euclidean distance between two different probes and distracters (A08 – A03; A09 – A10) was lower than between the two images of the same person (probe and target), meaning that if using these vectors it would not be possible to confidently state that the distracter and probe were not the same person. Interestingly, Actor 09 and Actor 10 had also been matched by seven different participants in the pilot similarity study, which was within the top 1% of highest matches, denoting an extremely high degree of facial similarity within this homogenous database.

Furthermore, these particular actors had also been paired together by separate pilot participants for the construction of matched pairs for the first live actor study (Experiment 6.1). In trials involving these particular actors in that experiment, 29.3% of participants in target absent trials had wrongly misidentified them as the target shown on video. Therefore, it is possible that one of these particular actors could be accused of a crime committed by the other, purely based on similarity of appearance. To prove innocence, the defence might engage an expert witness. However, the use of these limited permanent distance measurements alone could not reliably discriminate between the two different faces of these actors.

#### 10. 6. 2. 3. 2. Matching permanent anterior distance and angle vectors

The second hierarchical cluster analysis was conducted with the added inclusion of the angular measurements and the results are presented in Table 10.3. In this example, all probes and targets have lower distances than any probe has with a distracter, demonstrating that the addition of angular data provides greater discriminatory power.

*Table 10.3: Permanent anterior distance and angle measurements (Appendix C.2.1, C.2.2 & C.2.3)*

| Squared Euclidean | | | | Probe names | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance from probe | A01 | A02 | A03 | A04 | A08 | A09 | A10 | A33 |
| Target | 9.26 | 6.97 | 1.97 | 3.77 | 11.23 | 8.37 | 6.72 | 6.05 |
| Closest distracter | 12.71 | 16.20 | 15.81 | 15.45 | 13.03 | 10.18 | 9.89 | 20.59 |
| Distracter ID code | A88 | A42 | A55 | A92 | A03 | A11 | A11 | A23 |
| Similarity matches | 2 | 0 | 0 | 0 | 4 | 5 | 3 | 2 |
| Lowest distance between any | | 6.53 | A27 | Highest distance | | | 240.6 | A64 |
| two faces | | (0) | A89 | (similarity matches) | | | (0) | A76 |
| Overall database mean (SD) | | 60.10 (31.39) | | | | | | |

However, the lowest distance between two different faces within the database (6.53; A27 & A89) is still less than between six probes and their target faces. Interestingly, none of the 70 facial similarity matrix pilot participants matched these two faces, indicating that perceptual similarity judgements are not necessarily related to facial structure. Indeed, these particular faces would probably not be mistaken for one another. However, these findings indicate that even with the addition of the angular measurements, identity-matching using *DigitalFace* would still not be forensically reliable.

**10. 6. 2. 3. 3. Matching permanent and transient distance measurements only**

The third anterior probe analysis examined all permanent and transient distance vectors, while omitting the angular vectors (Table 10.4). In a forensic photo-anthropological assessment, care would be required before including transient features as some would be more stable than others. With very short hair, the vertex and trichion would be reliably identified, whereas, in contrast, eyebrows can normally be deliberately altered.

In this analysis, the distance between all probes and targets was less than the distance between any probe and its closest distracter, demonstrating that this set of measurements provides more item discriminatory power. Again though, the lowest distance between two different faces (5.76; A05 & A17) was less than that between five of the probes and targets.

*Table 10.4: Anterior permanent and transient distance measurements only (Appendix C.2.1 – C.2.3)*

| Squared Euclidean | Probe names | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance from probe | A01 | A02 | A03 | A04 | A08 | A09 | A10 | A33 |
| Target | 13.09 | 5.99 | 4.14 | 3.43 | 10.02 | 10.22 | 8.14 | 2.52 |
| Closest distracter | 16.22 | 11.20 | 13.29 | 10.82 | 13.33 | 11.56 | 9.28 | 18.07 |
| Distracter ID code | A19 | A3 | A11 | A28 | A57 | A57 | A78 | A76 |
| Similarity matches | 11 | 13 | 7 | 0 | 0 | 3 | 1 | 1 |
| Lowest distance between any two faces | | 5.76 (4) | A05 A17 | Highest distance (similarity matches) | | | 204.8 (0) | A96 A98 |
| Overall database mean (SD) | | 51.70 (26.30) | | | | | | |

In addition, some of the probes and distracters had similarity ratings in the top 1% of matches, meaning that these could potentially be mistaken for one another by humans (A01 & A19; A03 & A11; A02 & A03). Furthermore, *DigitalFace* identified Actor A02 as having highly similar measurements to A03, and with 13 matches within the subjective similarity database, this was the pairing rated tenth most similar. The video and photographic stimuli for these two actors (A02 & A03) were pooled together in Experiment 4.1 in which participants made decisions to arrays of six faces. When A02 was depicted in the video in target present conditions; of the participants who actually selected a face from an array, more chose A03 (23.9%), than selected the correct target (19.6%). A further 11.4% chose A03 in target absent conditions.

Similarly high error rates to this particular pairing were found in the same experiment when A03 was depicted in video. Although more participants who made a positive identification in target present conditions correctly identified A06 (50.9%), a large minority identified A02 as being in the video (31.4%) and in target absent conditions, 33.3% of all positive selections were again of Actor A02.

Although *DigitalFace* successfully discriminated between these two actors in this analysis, care would need to be taken if this had been part of a forensic investigation.

#### 10. 6. 2. 3. 4. Matching all anterior distance and angle vectors

The results of the final anterior view hierarchical cluster analysis with the inclusion of all distance and angular vectors are presented in Table 10.5. The squared Euclidean distances between all probes and targets were again less than between any probe and its closest distracter. However, the lowest distance between two different faces (15.00; A05 & A17) was still less than that between two of the probes and targets (A01 & A08), meaning that even with all anterior measurements included in the database, caution would be required before concluding these pairs of images actually displayed the same person.

### 10. 6. 2. 4. Matching profile probe undisguised faces with database targets

Two separate cluster analyses were conducted using profile view probe images of the 8 actors taken while attending identification sessions for Experiment 6.1. The first examined permanent distances only. The second included all angular and transient measurements.

*Table 10.5: Anterior distance and angular measurements (Appendix C.2.1 – C.2.4)*

| Squared Euclidean | | | | Probe names | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance from probe | A01 | A02 | A03 | A04 | A08 | A09 | A10 | A33 |
| Target | 16.30 | 11.63 | 5.38 | 5.32 | 15.53 | 13.72 | 12.02 | 6.53 |
| Closest distracter | 24.29 | 25.49 | 22.35 | 23.91 | 23.53 | 18.35 | 18.61 | 27.47 |
| Distracter ID code | A54 | A03 | A11 | A10 | A07 | A76 | A11 | A74 |
| Similarity matches | 0 | 13 | 7 | 7 | 1 | 0 | 3 | 1 |
| Lowest distance between any | | 15.00 | A05 | Highest distance | | | 298.3 | A40 |
| two faces | | (4) | A17 | (similarity matches) | | | (0) | A96 |
| Overall database mean (SD) | | 81.10 (38.48) | | | | | | |

#### 10. 6. 2. 4. 1. Matching permanent undisguised profile vectors only

*Table 10.6: Profile permanent distance measurements only (Appendix C.2.5 & C.2.6)*

| Squared Euclidean | | | | Probe names | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance from probe | A01 | A02 | A03 | A04 | A08 | A09 | A10 | A33 |
| Target | 2.59 | 1.93 | 3.04 | 2.07 | 2.51 | 3.19 | 2.47 | 2.23 |
| Closest distracter | 10.70 | 5.42 | 6.58 | 3.96 | 9.24 | 5.80 | 6.51 | 8.78 |
| Distracter ID code | A20 | A18 | A70 | A11 | A07 | A35 | A40 | A27 |
| Similarity matches | 3 | 2 | 0 | 1 | 1 | 1 | 2 | 4 |
| Lowest distance between any | | 2.54 | A23 | Highest distance | | | 89.05 | A12 |
| two faces | | (0) | A28 | (similarity matches) | | | (1) | A72 |
| Overall database mean (SD) | | 23.82 (11.96) | | | | | | |

The first profile cluster analysis examined the ten permanent profile distance measurements only (Table 10.6). All squared Euclidean distances between probes and targets were lower than between probes and distracters, passing one of the criteria for accepting a match. However, the lowest distance between two different faces on the database (2.54; A23 & A28) was again less than that for three of the targets with their probes meaning caution would be required before any conclusions were made that these images depicted the same person (A01; A03; A09).

#### 10. 6. 2. 4. 2. Matching all profile vectors

The second cluster analysis conducted on the profile images included all measurements in the set (Table 10.7). On this analysis, the squared Euclidean distance between all probes and their respective target images was less than that between probes and distracters, indicating a high degree of discrimination power. However, even with the inclusion of all profile vectors the lowest distance between two faces on the database (4.50; A23 & A28) was less than between five of the probes and their targets. This indicates that the use of all these potential profile vectors would not be sufficient to pass all criteria that would indicate a reliable match.

*Table 10.7: All profile measurements (Appendix C.2.5 – C.2.8)*

| Squared Euclidean | | | | Probe names | | | | |
|---|---|---|---|---|---|---|---|---|
| Distance from probe | A01 | A02 | A03 | A04 | A08 | A09 | A10 | A33 |
| Target | 14.02 | 3.40 | 6.46 | 3.54 | 5.89 | 4.60 | 4.49 | 5.71 |
| Closest distracter | 18.62 | 23.27 | 14.41 | 12.42 | 17.77 | 18.97 | 9.21 | 15.58 |
| Distracter ID code | A53 | A27 | A11 | A57 | A07 | A35 | A40 | A64 |
| Similarity matches | 0 | 1 | 7 | 0 | 1 | 1 | 2 | 1 |
| Lowest distance between any | | 4.50 | A23 | Highest distance | | | 165.9 | A02 |
| two faces | | (0) | A28 | (similarity matches) | | | (0) | A61 |
| Overall database mean (SD) | | 45.28 (20.38) | | | | | | |

#### 10. 6. 2. 5. Matching restricted vectors due to disguise

The final cluster analysis examined the effectiveness of the *DigitalFace* system to correctly match and discriminate faces using a restricted database of measurements due to the use of a disguise (Table 10.8). For this the database containing the 70 anterior and profile view images was utilised. The probe images were those obtained of the eight actors wearing the hat, photographed on the same day as their undisguised database images. A list of the restricted 19 distance and 6 angular vectors in anterior view and 7 distance and 8 angular vectors in profile view is presented in Appendix C.2.9.

In this example, the squared Euclidean distance between all probes and their specific targets was less than the distance between probes and distracters. Furthermore, the lowest distance between two different images on the database (12.70; A23 & A39) was more than that between any probe and its respective target, indicating that the use of both profile and angular measurements, even when restricted by the use of disguise is a more reliable technique than using profile or anterior measurements alone.

*Table 10.8: Frontal and profile restricted permanent distance and angles (Appendix C.2.9)*

| Squared Euclidean Distance from probe | Probe names A01 | A02 | A03 | A04 | A08 | A09 | A10 | A33 |
|---|---|---|---|---|---|---|---|---|
| Target | 11.20 | 5.19 | 11.45 | 3.21 | 6.84 | 4.73 | 5.74 | 6.62 |
| Closest distracter | 30.67 | 23.51 | 15.53 | 15.89 | 20.09 | 29.02 | 11.78 | 30.03 |
| Distracter ID code | A20 | A27 | A07 | A10 | A07 | A03 | A17 | A15 |
| Similarity matches | 3 | 1 | 4 | 7 | 1 | 8 | 3 | 1 |
| Lowest distance between any two faces | | 12.70 (4) | A23 A39 | Highest distance (similarity matches) | | | 206.5 (0) | A69 A70 |
| Overall database mean (SD) | | 62.49 (26.71) | | | | | | |

### 10. 6. 2. 6. Disambiguation of misidentified faces in previous experiments

If *reliable* photo-anthropometric measurement differences are found between two images it can safely be stated that they are not of the same person. Indeed, it is far easier to discriminate between two different people than it is to decide if two images of the same person are matched. Therefore, the final series of analyses conducted using the *DigitalFace* system was designed to examine whether it would successfully discriminate between faces that had been wrongly identified most often as the same person in each of the experiments reported in this thesis. The errors associated with each pair can be found in relevant chapters.

*Table 10.9: Squared Euclidean distance and rank on anterior, profile and combined databases for each pair of faces associated with the highest errors in some of the experiments in this thesis*

| | Anterior | | | Profile | | |
|---|---|---|---|---|---|---|
| | Squared Euclidean Distance | Rank Item | Rank Database | Squared Euclidean Distance | Rank Item | Rank Database |
| Pilot study 3.7 (subjective similarity study – most paired items in database) | | | | | | |
| A29 & A30 | 55.10 | 29 | 1,592 | 49.46 | 30 | 1,500 |
| Experiment 4.1 (Section 4.1.2.3; six alternative choice matching study) | | | | | | |
| Target present conditions | | | | | | |
| Video: A32* Photo: A 31 | 81.54 | 76 | 3,260 | 49.63 | 39 | 1,667 |
| Target absent conditions | | | | | | |
| Video: A32: Photo: A25 | 55.00 | 53 | 1,582 | 15.44 | 3 | 65 |
| Experiment 4.2 (Section 4.2.2.4; single-item identity-verification design) | | | | | | |
| Video: A24: Photo: A25 | 74.87 | 52 | 2,907 | 25.95 | 6 | 364 |
| Experiment 5.1 (Section 5.1.2.5; single-item identity-verification design – no disguise in video) | | | | | | |
| Video: A37**: Photo: A36 | 30.85 | 1 | 158 | 21.49 | 7 | 203 |
| Experiment 6.1 (Section 6.1.2.3; single-item live identity-verification design) | | | | | | |
| Video: A01*** Live: A02 | 33.45 | 6 | 258 | 78.47 | 63 | 2,178 |
| Experiment 6.2 (Section 6.2; single-item live identity-verification design – footage 1-year old ) | | | | | | |
| Video: A04: Live: A03 | 35.71 | 12 | 345 | 27.71 | 16 | 460 |
| Experiment 7.1 (Section 7.1.3.1; single-item live identity-verification design – close-up footage) | | | | | | |
| Video: A45 Live: A46 | 40.37 | 1 | 590 | N/A | N/A | |

**A31 & A32 (also subject of most errors: Experiment 5.2; hat condition)*

***A36 & A37 (also subject of most errors: Experiment 5.1; glasses, hat & Experiment 5.2 - no disguise conditions)*

**** A01 & A02 (also subject of most errors: Experiment 5.2; glasses condition)*

Table 10.9 lists the squared Euclidean distance between these particular actor pairs on the anterior and profile proximity matrices for all measurements (permanent, transient and angular). It would be possible using this information in court, to state whether other individuals within the population would be likely to possess similar facial vectors. Therefore, firstly, the *item* rank of each pairing is provided. This denotes the number of faces within the database possessing the same or a lower squared Euclidean distance to that specific target face, than that between the two misidentified faces. The *database* rank indicates the number of pairings within the entire database possessing the same or a lower squared Euclidean distance.

Two of the pairs of faces, A36 and A37 (Experiments 5.1 & 5.2); and A45 and A46 (Experiment 7.1) were both the subject of the most misidentifications in these respective experiments and were also first on the anterior view item rank, indicating that no other face was closer in Euclidean space on the *DigitalFace* measurement data. However, many other pairs of faces in the database possessed lower Euclidean distances. This illustrates that by specifying a minimum criterion for a match to be lower than any pair within the database provides increased safety against making a wrong decision, than would be a decision rule based on whether any one face on the database possessed a lower Euclidean distance to the target.

### 10. 6. 3. General discussion

The findings of the photo-anthropometrical analyses conducted in this chapter illustrate that great caution should be taken when applying this methodology if attempting to determine that two different photographic images depict the *same* person. In contrast, in the majority of trials, the application of the technique correctly disambiguated many of the pairs of *different* faces in good quality images that had been the subject of the highest number of identification errors in experiments reported in this thesis. Thus, the use of *DigitalFace* was often more reliable than observations made by humans. Although viewpoint in probe and target images of the same person did not exactly match, the quality of the photographic images used in these analyses allowed for the optimal measurement of a large number of distance and angular vectors in both anterior and profile views. It is unlikely that images obtained as part of a criminal investigation would afford such fine detail. Moreover, expert witnesses would probably only be asked to apply their techniques when images were impoverished in some manner, or if the appearance of the defendant had changed, for instance, by growing a beard. Indeed, under UK law, an expert should only be called to present evidence if a jury would be unlikely to be able to form an opinion without that assistance (e.g., *R v Turner*, 1975). If images were of such high-quality as those used in this chapter, a jury would almost certainly be invited to make their own unaided visual comparison (Attorney General's Reference, No. 2 of 2002, 2003).

With low-resolution or unclear images, particularly if the subject is sited some distance from the camera, or viewpoint angle is not exactly matched, the identification of landmarks would be more problematic, limiting the number of measurements that could be denoted and increasing the likelihood of error. These findings illustrated that even when *all* vectors were included in anterior view, and separately with all vectors in profile view, the squared Euclidean distance between two images of two different people was sometimes less than that between two images of the same person. Thus, using a limited number of facial probes against this limited database, it was not possible using this photo-anthropometrical methodology to correctly categorise these particular images. A reduction in the number of landmark sites, or an increase in the inaccuracy of landmark identification could seriously increase the error rate. Only when anterior *and* profile viewpoint data was included *together* in an analysis, were all probe images correctly categorised. This was actually found using the restricted set of vectors due to probes being in disguise. Because of this, there was no necessity to conduct a further evaluation with the inclusion of *all* anterior and profile vectors from undisguised faces.

The use of high-quality images proved to be an ideal test for appraising the *DigitalFace* system, as the analyses highlighted difficulties involved in photo-anthropometry due to the highly similar structure of the human face, even with a full set of vectors. The results of the error determination analysis demonstrated that if all vectors were misidentified by only one standard deviation from the mean, this distance in Euclidean space was equal to 2.25 units. And yet, prior data screening to remove the influence of outliers more than 2 SD from the mean had already been performed. If that margin of error is applied to all the analyses reported in this chapter (and in a forensic case were to be applied to images of poorer quality), a number of classifications that had correctly disambiguated image pairs would require caution. However, this provided an estimation of the error associated with landmark identification using a single image. It is possible that in a forensic case, with moving images, multiple photographic frames may be acquired meaning a more robust error analysis could be conducted by comparing measurements across frames. From this, confidence levels could be presented along with an opinion concerning identity. In addition, inter-rater reliability could be assessed by the employment of more than one landmark marker.

There have been repeated calls for the establishment of large-scale databases of facial measurements in order to assess the safety of identification matching using facial mapping techniques. Indeed, without a:

> "National database of facial characteristics, or any accepted mathematical formula, as in the case of fingerprint comparisons – any expression of the *degree of support* provided by particular facial characteristics or combinations of facial characteristics must be only the *subjective* opinion of the facial imaging or mapping witness" (*R v Gray*, 2003, p. 3).

The anterior database for the analyses reported in this chapter contained 100 targets and distracters, and the profile database contained 70 faces. It could be argued that this was not sufficient to fully evaluate the technique. However, unlike some previous facial mapping research (Vanezis et al., 1996; Yoshino et al., 2000; 2002), the homogenous inclusion criteria ensured that the distracter faces were representative of the population being tested. In addition, the technique provided an indication of the commonality of facial proportions, among even this relatively limited database. Nevertheless, an increase in the number of faces included in the database meeting these criteria would probably result in more faces possessing similar facial dimensions, which would increase the potential of error in matching identity from two different images. It would also be necessary to create further facial databases, if, for instance, the system was to be forensically applied to those of different ethnic backgrounds and age ranges or female targets.

Reliable differences between the measurements of different faces in images, particularly, those wrongly matched by human participants meant it was possible to confidently state that they did not depict the same person. However, there are two circumstances in which differences in measurements would not necessary indicate different people. The first would be if camera viewpoint was not matched in both images. Secondly, there have been attempts by analysts to compare images taken decades apart (Iscan, 1993). However, ageing is accompanied by a number of specific changes to facial structure, including consistent growth of the jaws and nose throughout childhood, which alters the position and relative size of the eyes. This heart-like expansion of the head from a constrained nodal point at the junction of the brainstem and spinal cord has been described using a mathematical approximation called *cardioidal strain* (Bruce, Burton, Doyle and Dench, 1989). Other changes occur; - cartilage continually grows in the nose and ears, tooth loss can make noses and chins more pronounced and traumas can add scars. Fat and muscle tissue relaxes and the skin loses its elasticity, causing wrinkling and sagging (George & Hole, 1995). Changes to hair and skin colour, and perhaps more importantly the receding and thinning of hair all alter appearance as can cosmetic changes such as plastic surgery, mole removal, hair colouring and make up. Finally, the first lip, nose and chin transplant was recently performed (Lichfield, 2006). Theoretically, a criminal determined to evade conviction could radically change their perceived appearance by 'facial transplant'. Although an extreme situation, this would make it virtually impossible to apply any of the photographic comparison techniques.

### 10. 7. Summary

Photographic video superimposition, being based purely on human observation, even if demonstrated in public would never be sufficiently objective to meet standards of scientific rigour. Furthermore, morphological classification analysis, by definition, involves grading facial features into pre-determined discrete categories, which may not be flexible enough if a specific feature possesses elements of more than one category, or is on the boundary between two. Indeed, because all analyses must be conducted at a nominal level, it would be difficult to statistically discriminate between two different faces possessing similar characteristics. As such, the technique would be unlikely to distinguish between siblings of the same gender and probably more distant family members and even many of the same age and ethnic background. Without this level of assessment, tested against a database of faces, the Court of Appeal ruled that any intimation of a quantified measure of support for a match could not be objectively evaluated (*R v Gray*, 2003). However, perhaps surprisingly, more recently, the same court has ruled that knowing the likelihood of shared facial characteristics is not necessary when providing this type of evidence as expert testimony is admissible even if only a *subjective opinion* is being expressed (*R v Gardner*, 2004).

Nevertheless, a review of the manner in which expert evidence is admissible in the UK has been ongoing since the autumn of 2005, mainly due to a number of medical cases in which scientific evidence was found to be questionable (BBC News, 10 October 2005). It is possible that this review may recommend the adoption of current safeguards in the USA, in order to curtail against experts providing subjective judgements based on 'junk science' (*Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 1993). As such, the methodology is required to have gained general acceptance in its particular academic discipline, to have been scientifically tested, subjected to publication or peer review and it should be accompanied by the calculation of error rates (real or potential; Groscup, Penrod, Studebaker, Huss & O'Neil, 2002). The studies conducted in this chapter have demonstrated that photo-anthropometric analysis surpasses other photographic comparison methodologies in terms of validity and reliability. It was also found to be more reliable than many human participants. The ability to produce an estimation of error and when used in conjunction with a database of suitable faces, to provide a quantifiable likelihood that two images are, or are not, of the same person would provide robust evidence.

Indeed, if accepted for peer-reviewed publication, the methodology would pass all *Daubert* criteria. Furthermore, given sufficient photographic quality, the preciseness of measurement could potentially even distinguish between monozygotic twins.

## Chapter 11: General discussion and conclusions

### 11. 0. Introduction

There are currently millions of surveillance cameras throughout the world and these numbers are certain to substantially increase in the next few years (Norris & Armstrong, 1999; Norris et al., 2004). Images are often used by the criminal justice system for identification purposes and the growth in CCTV coverage is likely to result in additional court cases in which evidence of this type forms the basis of the prosecution case. Although suspects often confess to crimes when they are informed that CCTV evidence depicts an offence, a number of legal principles have been established in the UK as to the manner in which photographic and video identification evidence may be presented in court when identification is disputed (e.g., Attorney General's Reference, No. 2 of 2002, 2003). This review of the admissibility of CCTV evidence in court concluded that:

> "In our judgement, on the authorities, there are, as it seems to us at least four circumstances in which, subject to the judicial discretion to exclude, evidence is admissible to show and, subject to appropriate directions in the summing up, a jury can be invited to conclude that the defendant committed the offence on the basis of a photographic image from the scene of the crime" (Attorney General's Reference, No. 2 of 2002, 2003).

The experiments reported in this thesis were designed to investigate some of these principles, as well as other procedures that might be adopted by the criminal justice system when the identity of an individual shown in video evidence is in dispute.

### 11. 1. Familiar face identifications

The first of the four circumstances relates to testimony given by witnesses claiming to know "the defendant sufficiently well to recognise him as the offender depicted in the photographic image" (Attorney General's Reference, No. 2 of 2002, 2003). Such a witness may give evidence as though a bystander at the actual incident and in summing up the judge would normally refer to the potential for error given the specific circumstances of the case.

Empirical evidence suggests that if a person is highly familiar with an individual they will be accurate at recognising that person or in deciding that they are not present, even in extremely poor-quality video or stills (e.g., Bruce et al., 2001; Burton et al., 1999). Indeed, face recognition can be extremely good when sufficiently familiar, as it is possible to distinguish between monozygotic twins (Rhodes, 1988). In the pilot study conducted as part of this thesis, acquaintances of actors depicted in the medium-range simulated CCTV footage used in a number of experiments were 100% correct in identifying their friends (Section 3.4[1]). In contrast, people unfamiliar with the same actors made many identification errors to the same images across different experiments in this thesis. These findings add credence to the weight that should be placed on evidence of this type in court, suggesting that as long as a witness can demonstrate a high level of familiarity with the defendant, their testimony is likely to be reliable.

However, it is unclear how much experience is required before an individual would be considered 'highly familiar' in order that recognition of facial images would be reliable. For this reason, identification evidence from CCTV by police officers is not normally admissible in Australia. It is considered that a jury by the time of deliberation is likely to have been exposed to the defendant for a longer period than the past experience of most police officers (*Smith v The Queen,* 2001).

### 11. 2. Unfamiliar face identifications

The second circumstance listed in the Attorney General's Reference (Attorney General's Reference, No. 2 of 2002, 2003) was that "where the photographic image is sufficiently clear, the jury can compare it with the defendant sitting in the dock".

[1] In addition, some visitors to the Science Museum ($n$ = 8), volunteering to participate in experiments viewed the images and recognised and accurately named some of the depicted actors. None knew the experimenter and one particular individual coincidently from the same university was able to name five of the six actors. Their data was not included in analyses.

It is also expected that a contemporary undisputed photograph of the defendant would be available for inspection. Members of a jury will always be unfamiliar with the defendant and in contrast to the impressive recognition rates associated with familiar faces, unless exactly the same images are used in learning and test (e.g., Bruce, 1982), the identification of unfamiliar people in even the highest-quality images is error prone. This is even found in matching studies with no demands on memory (e.g., Bruce et al., 1999; Bruce et al., 2001; Davies & Thasen, 2000; Henderson et al., 2001).

However, in all these simultaneous matching experiments, participants have been asked to make decisions from moving video footage or stills to photographs. In court, the defendant will always be present in person, meaning more identification cues would potentially be available. Therefore, experiments in this thesis were conducted to examine the reliability of decisions in this context. Additional experiments were also conducted in which photographs rather than live actors were targets. These were designed to firstly examine whether the results of previous published research would be replicated and secondly, whether jurors might be helped to make a correct identification by being presented with a contemporary photograph of the defendant.

A further issue in three of the experiments was to examine age group differences in simultaneous matching ability, particularly the performance of children. Whilst children would never act as jurors in court, the findings of these perceptual comparison studies are of theoretical interest and have implications in terms of the reliability of eyewitnesses to identify an offender.

### 11. 2. 1. Age group differences in unfamiliar face matching to video

In the first experiment in Chapter 4, participants viewed target actors shown on video and attempted to match each target to their photograph within an array of six. The targets were absent from the arrays in two of the six trials. The videos simulated medium-range CCTV footage and the facial detail available was probably higher than in most evidence depicting actual criminal incidents. Comparison photographs were close-up three-quarter view portraits taken within 24 hours of the video footage. The main findings were that young children (8 – 9 years) and adolescents (12 – 14 years) were less accurate than young adults (18 – 23 years) at identifying the target actors in the arrays. These results supported previous research finding a developmental improvement throughout childhood until approximately 11 years-of-age followed by a dip in face processing ability at the start of adolescence (e.g., Carey et al., 1980; Chung & Thomson, 1995; Diamond et al., 1983; Soppe, 1986). Indeed, the performance profile of the 12 – 14 year-olds was closer to the 8 – 9 year-olds than to the immediately younger 10 – 11 year-old age group.

Older adults aged 45 - 76 were also slightly worse than the young adults at this task. The target actors were all from the same age group as the young adult participants and previous studies have found a face processing and recognition advantage for faces of a similar age as the viewer (e.g., George & Hole, 1995; Wright & Stroud, 2002). However, none appear to have previously found differences in matching performance across adult age groups.

Nevertheless, in contrast to the results summarised above, no age group differences in identification accuracy were found in the second experiment reported in Chapter 4, using a single-item identity-verification task. In half of the six trials the target was present in the video; in the remainder a distracter was present. It could be suggested that due to task simplicity, reliable age group differences were not detected. Indeed, with nearly 500 participants, the design possessed potentially high statistical power. However, performance varied *within* all age groups and although many performed at ceiling (13.9%), this was not restricted to young adults. If the results of these experiments are considered together it suggests that age group deficits occur when attempts are made to distinguish between groups of similar appearing faces from a different age group to the participant. However, when only *one* face needs to be matched, there is no interference from other distracters and therefore there is no additional detriment to performance.

The third experiment to contrast the performance of adults and children employed the same single-item identity-verification design. However, instead of using photographs, the target actor was present in person. The main age group related findings revealed that a higher proportion of children than adults responded that the live actor was present in the video, regardless of whether this was true or not. This resulted in an increase in the number of target present *correct* selections, while having the contrasting effect of increasing the number of *incorrect* target absent selections.

Conclusions from this experiment are limited as only one actor was recruited, meaning results might not generalise to the wider population. Furthermore, the video images sometimes depicted both the target actor (in target present trials) and the distracter (in target absent trials) in disguise and were taken a year prior to the identification sessions, increasing the difficulty of the task. Half of the participants were warned of this although a warning equally assisted adults and children in decision-making and indeed, no differences were observed between age groups due to the use of disguise. Nevertheless, this experiment is possibly the first to demonstrate differences in simultaneous matching ability between children and adults when the target is present in person.

### 11. 2. 2. Unfamiliar face matching by adults

In terms of decisions that may be required by a jury, of forensic interest are the results relating to adults only, as 18 is the minimum age for jury service in the UK. The two experiments reported in Chapter 4 directly compared the performance of older adults and younger adults. In the first, older adults selected slightly more incorrect distracters from arrays than younger adults. No differences in the number of errors made by participants of different ages were found in the second experiment, although older adults did tend to respond with higher confidence when making both correct and incorrect decisions.

From these findings a question mark might be raised if a predominately older group of jurors were presented with CCTV evidence depicting a young adult for them to compare the resemblance with a defendant of that age. These results suggest that they may be more likely to confidently respond that identity matched, regardless of whether this was correct. However, due to the recruitment of proportionately fewer older adults in the remaining experiments in this thesis, it was not possible to further examine adult age group performance differences.

In the first experiment in Chapter 4, when targets were present in arrays of six photographs, the correct hit rate for all adult participants, regardless of age group was 53%, a further 30% were incorrect false positive responses and the remainder involved incorrect 'not present' responses. In target absent conditions, approximately 64% of trials involved the incorrect selection of a distracter. These results confirm the inherent difficulties found in similar previous research examining matching to photographs in arrays when all targets are unfamiliar (Bruce et al., 1999; Davies & Thasen, 2000; Henderson et al., 2001). However, a jury would only be required to make these matching decisions based on whether a single defendant was the offender pictured in video footage. The second experiment in Chapter 4 was designed to replicate this task; however, rather than an actor being physically present a contemporary photograph was available. Even with these reduced task demands, in target present conditions approximately 29% of trials involved an incorrect belief by participants that the target shown in video was not depicted in the photograph. The proportion of errors was higher in target absent conditions (32%), signifying an incorrect belief that the same person was depicted in both images. As reported in Chapter 3, pilot participants were 100% accurate at recognising their acquaintances in the same images, demonstrating that the problem is not due to image quality, but is related to familiarity.

Imprisoned offenders have stated that they would be more likely to wear a disguise to avoid detection from CCTV (Loveday & Gill, 2003). The two experiments in Chapter 5 were designed to address this issue as the target actors were filmed and photographed in three disguise conditions (no disguise, dark glasses or a hat). The single-item identity-verification design was again employed. In the first experiment, the videos depicted the actors in one of the three disguise conditions, whereas in the photographs they were all shown in no disguise. In contrast, in the second experiment, disguise was varied in the photographs with all videos depicting the actors in no disguise.

In both experiments, in target absent conditions, approximately 30% of responses were incorrect. This was consistent, regardless of whether the images were depicted in disguise or not. In contrast, when targets were present in both images, the most errors were made when one image depicted an actor wearing a hat, followed by when they were wearing dark glasses and the least when they were shown in no disguise. These results are consistent with laboratory studies finding that the removal or the addition of a disguise reduces later recognition accuracy (e.g., Diamond & Carey, 1977; Hockley et al., 1999). However, a reduction in the number of correct hits is normally accompanied by an increase in false alarms. In these matching experiments, disguise affected the target present hit rate only. Confidence in decisions was also associated with disguise in both experiments, in that it was highest when actors were shown in no disguise in either image, and lowest when wearing a hat.

Across the two experiments 1,200 participants were recruited, providing statistically powerful data. These findings suggest that if CCTV images were obtained of a crime scene and the offender was depicted wearing a disguise, a jury asked to compare the image with the defendant would be more cautious in their responses. This would tend to favor the defendant as it would potentially increase the likelihood of a not guilty verdict. However, in contrast, the use of one of these disguises would not additionally appear to increase the likelihood of a wrongful conviction involving an innocent defendant.

The results of the experiments summarised so far illustrate that the matching of a video image with a contemporary high-quality photograph is error prone. And yet, this practice is recommended if a jury is invited to examine surveillance images of an offender to compare the resemblance with the defendant (Attorney General's Reference, No. 2 of 2002, 2003). However, the defendant would obviously be present in person in court and additional identification cues may be available to improve decisions. The two adult-participant-only experiments reported in Chapter 6 were designed to examine these circumstances, using the same single-item identity-decision design and medium-range video images as before. However, instead of targets being depicted in photographs, actors, all of whom were unfamiliar to participants, attended live identification sessions in person.

In the first experiment in Chapter 6, eight white male actors aged 18-21, were employed to act the part of 'defendants' and 'culprits' with the identification sessions taking place approximately 3 weeks after filming the video footage. During each session, the 'defendant' walked into the room and stood by a television screen while the footage, depicting the 'culprit' in no disguise was playing. Identity-verification was again found to be fallible and in many of the individual actor sessions, performance was at chance levels. Across all target present trials, 22% of the participants wrongly believed that defendants were not depicted in video. In target absent trials, approximately 17% incorrectly identified defendants present in person as being in the videos. However, values varied across trials involving different defendants. In one case, 44% of participants incorrectly judged an 'innocent' defendant as being the culprit in video. In another, 37% wrongly responded that a defendant was not the culprit shown in the footage. In other cases, performance was 100% accurate.

Most criminal cases take far longer than three weeks to reach the courts, especially those appearing in crown court in which a jury will be required to reach a verdict. To examine this in the second experiment reported in Chapter 6, two of the defendants from the first experiment were again recruited; attending identification sessions a year later. Half of the 591 participants were correctly informed in advance that the video footage was a year old. The videos depicting the actors wearing a hat and dark glasses were also used to measure the effect of disguise on matching to live actors. Overall error rates were higher than in the first live actor experiment of Chapter 6. In target present conditions 44% of participants made incorrect decisions, an indication that the appearance of the actors had changed in the time interval. In target absent trials 33% of decisions were in error.

The warning concerning the age of the video was found to increase decision accuracy across all conditions involving one defendant, while decreasing accuracy with the other, suggesting that caution is required before generalising results from these single actor studies to the wider population. Furthermore, the overall effects of disguise were not consistent with those found using photographs in Chapter 5. Indeed, accuracy was highest when actors were depicted in dark glasses with no differences between error rates in the other two conditions.

A warning as to the age of the video had no effect on error rates in the no disguise and in the hat conditions. In contrast, when the video depicted the culprit wearing a pair of dark glasses, a warning was found to slightly increase the number of errors in target absent conditions while reducing them when targets were present. These findings illustrate that in some circumstances, knowledge that a video is not of recent origin may increase willingness to accept that minor differences in appearance are due to the effects of time. As all jurors would be aware of the date of the crime for which a defendant was being tried, these results suggest that this might increase the likelihood of a miscarriage of justice if an innocent suspect had been charged.

From the results of the experiments reported in Chapters 4, 5 and 6 using footage that might be captured by typical open-street CCTV systems, it was apparent that the high error rates must be partly due to the difficulties experienced in attempting to distinguish facial features in the medium-range video images. Therefore, the experiment reported in Chapter 7 was designed to examine live actor face matching using close-up facial images continually turning from left-to-right profile. In half the trials a full-face photograph of the defendant was presented, to directly compare performance with trials attended by a physically present actor.

Performance was again found to be error prone in both target present and target absent trials. The fewest errors (16.7%) occurred when target present identification sessions took place a few minutes after the video footage was filmed, regardless of whether defendants were shown in photographs or were physically present. However, errors were higher in a second target present condition using footage taken a week previously, with actors having changed their appearance slightly. Approximately 26% of participants made incorrect decisions in this condition when actors were live. However, possibly due to participants having fewer identification cues when actors were depicted in a photograph, the error rate was higher in this condition (40.3%).

There were also high error rates when distracters were depicted in the videos in target absent conditions. However, in contrast to the target present condition the pattern of errors were reversed. More errors were made when the defendants were physically present (41.3%) than when depicted in photographs (27.0%). These findings illustrate that even using high-quality video footage, human ability at what appears to be a relatively simple task is error prone. If replicated in a court room by members of a jury, the consequences of this type of error could be serious.

A further issue in relation to these findings is that in the UK, since the passing of the Crime and Disorder Act (1998), individuals can be issued with local authority assigned anti-social behaviour orders (ASBO) based on a portfolio of evidence. Rather than requiring evidence of guilt 'beyond reasonable doubt' as in a criminal trial, they can be issued on the less stringent 'balance of probabilities' civil law burden of proof. This can include the use of 'hearsay' evidence and if the order is broken, offenders can be imprisoned for up to 5 years. By 2006, approximately

7,000 ASBOs had been issued, in particular to reduce the occurrence of anti-social behaviour (*Independent*, 10 January, 2006).

CCTV recordings are often acquired in order to assemble this type of evidence and indeed, some local authorities have provided cameras free of charge to householders suffering from regular minor criminal activity (e.g., Peterborough, *Peterborough Evening Telegraph*, 12 January 2005). With no requirement to prove an individual guilty beyond reasonable doubt, evidence from video would certainly be sufficient for an ASBO to be issued. Furthermore, in contrast to the protection offered to suspects by criminal law, once an order is applied, there is often no legal recourse to challenge the ruling.

ASBOs based on CCTV evidence would be most likely to be issued on the evidence of a local police officer or the victim, suggesting that identification errors would be less likely (G. Davies, Personal Communication, 7 July 2006). However, from the results of previous experiments, in particular Experiment 7.1, it is clear that mistaken identifications in even the highest-quality images can occur. Furthermore, the case of mistaken identification by family members described in Chapter 1.5 illustrates that even highly familiar people can sometimes be wrongly identified in high-quality images, possibly due to contextual information. In addition, Burton et al. (1999) found that police officers were no better than normal members of the public at recognising unfamiliar people in poor-quality CCTV footage. Nevertheless, only a breach of the provisions of an ASBO is a criminal offence and this must be proved beyond reasonable doubt in a normal court and this increased burden of proof should provide some safeguard against wrongful convictions.

Furthermore, in a crown court, a jury of twelve would have to consider a verdict, meaning that some protection would be afforded against individual decisions. This was examined in Chapter 9 in which a series of mock jury 'trials' were conducted. In these, the only 'evidence' presented was one of the medium-range surveillance videos. Participants in groups of 12 were required to deliberate as to whether the footage depicted the 'defendant' shown in a photograph. Each jury delivered a verdict in two counterbalanced trials, one when the target was present, one in which the video depicted a distracter. Individual private responses were also recorded pre- and post-deliberation.

The main findings were that some of the juries delivered a guilty verdict based on the strength of video evidence alone. Indeed, one jury jointly found a defendant guilty, when a different person was actually depicted in the video footage, partly due to the influence of one juror. If replicated in real life, an innocent person would have been convicted. Due to the artificial nature of this type of design, it is not possible to conclude that similar findings would occur with real trials. However, all participants were specifically warned of the necessity to deliver a 'beyond reasonable doubt' verdict and to treat the experiment as though the consequence of decisions would be binding.

It was not possible to perform jury-level statistical analyses to measure the likelihood of a guilty verdict in these circumstances. Nevertheless, consistent with previous research (e.g., Davis et al., 1975; Devine et al., 2001; MacCoun & Kerr, 1988), juries with a majority voting guilty at the start of the deliberation process were more likely to deliver a final guilty verdict. The opposite was true when an initial majority favored a not-guilty verdict. Comparison of pre- and post-deliberation individual private responses found that all jurors in their first trial generally moved towards voting for leniency regardless of target presence. If the second trial was target absent, there again tended to be a movement towards leniency. In contrast, jurors receiving the target present trial second tended to retain their individual private responses, supporting previous research finding that prior jury experience can sometimes influence verdicts in later trials (Dillehay & Nietzel, 1985; Kerr et al., 1982; Nagao & Davis, 1980).

These findings of juror experience influencing verdicts have implications in terms of the fairness of the criminal justice system and the presumption of innocence.

Furthermore, there is also an implicit assumption that a jury of twelve will form some protection against mistaken individual judgements. However, the joint decision making of these groups in this experiment did not safeguard against an incorrect decision, particularly when influential 'jurors' were able to lead the deliberation discussions.

### 11. 3. 'Ad-hoc' expert witness identification testimony

It is apparent from the results of the experiments summarised so far in this chapter, as well as from previous research (e.g., Bruce et al., 1999) that identification matching involving unfamiliar faces from video footage to either photographs or to live actors is potentially error prone. It is also not possible to rely on group decision making as carried out by 12-person juries to ensure that the collective opinion protects against incorrect judgements. However, the Attorney General's reference allows for a third principle for enabling a jury to come to a collective decision:

> "Where a witness who does not know the defendant spends substantial time viewing and analysing photographic images from the scene, thereby acquiring special knowledge which the jury does not have, he can give evidence of identification based on a comparison between those images and a reasonable contemporary photograph of the defendant, provided that the images and the photograph are available for the jury" (Attorney General's Reference, No. 2 of 2002, 2003).

Such a witness would be described as an *ad-hoc* expert as the extensive experience of viewing the images would allow them to provide opinion evidence to a jury. The experiment reported in Chapter 8 was designed to examine whether such intensive viewing and inspection would result in higher identification accuracy than the limited time scrutinising the same images as would be expected of a jury. Using the medium-range video footage and a series of comparison photographs presented in arrays, participants completed a series of matching decisions. No differences in final identification accuracy were found between a learning group viewing the twelve actors on video at least 25 times over a three-hour time period and a control group who watched the same footage a maximum of six times.

From these results, it was not possible to conclude that testimony from a police officer conducting a close analysis of video footage would be more reliable than the perception of a jury member. However, there would be many other factors involved in such a case. These would include the quality of the footage, the distinctiveness of the offender and the specific procedures conducted by a police officer. However, a small minority of the learning group participants expressed surprise at the end of the experiment when provided with performance feedback that their responses were incorrect. After viewing the footage many times they were required to decide which of two photographs depicted the individual in the footage. One was always the target; the second was a photograph of the distracter that had been chosen by that participant most often from arrays across previous trials. These participants appeared convinced that the feedback they were receiving as to the identity of the actual target was incorrect and that the distracter they selected was actually present in the video. It would therefore be likely that if these particular participants had followed these procedures prior to attending court, for instance if a police officer, they would present an extremely convincing and confident argument as to the guilt of an innocent defendant.

### 11. 4. Expert witness identification evidence

The final principle referred to by the Attorney General, allows for identification evidence provided by expert witnesses applying their professional skills in facial structure or photographic image comparison:

> "A suitably qualified expert with facial mapping skills can give opinion evidence of identification based on a comparison between images from the scene (whether expertly enhanced or not) and a reasonably contemporary photograph of the defendant, provided the images and the photograph are

> available for the jury", (Attorney General's Reference, No. 2 of 2002, 2003).

Details of the three main approaches to facial mapping used by expert witnesses in court; - morphological classification analysis, photograph image superimposition and photo-anthropometric comparison analysis were discussed in Chapter 10. All have received criticism for a lack of scientific rigour, particularly when this type of evidence is one of the principle prosecution planks against the defendant. This has sometimes been a justification for applications to appeal against conviction. Some have been successful, with in one, the court ruling that due to a lack of any estimation of whether other people in the population possess similar facial properties to the defendant, it is not possible to provide more than a *subjective* opinion of whether they are shown in video (*R v Gray*, 2003).

One of the methods used by expert witnesses in these circumstances, photo-anthropometrical analysis was examined in depth in Chapter 10. This technique involves the measurement of facial landmark distances and angles in a photograph and new software (*DigitalFace*) was designed to aid in the identification of landmarks and to automatically calculate and produce a set of facial dimensions. A series of exploratory analyses were conducted to examine whether the facial dimensions of probe and target images of the same person taken at different times were more alike than those of a homogeneous database of facial images of 100 people of the same gender, ethnicity and of approximately the same age. Using a database in this manner, the method provided some assessment of the likelihood of occurrence of people with similar facial characteristics. Furthermore, an estimation of the associated error in applying the manually-assisted landmark location software was calculated. If presented in court, and tested against an appropriate facial database it would be therefore possible to provide an *objective* opinion as to identity.

There were two criteria for a positive identification decision. The first was that the probe face should be closer to the target face in terms of Euclidean space than to any other face within the database on the particular set of measurements included in the analysis. To fail this criterion would either indicate that the probe and target images were of two different people, or, that the number of measurements included was not sufficient to discriminate between different faces. The second, more robust criterion was that the distance in Euclidean space between the probe and target should be less than that between all other pairs of images in this database. If this criterion was passed in a forensic case using a similar database, it would be possible to claim with a measure of probability that the defendant was depicted in the comparison image.

A number of separate analyses were conducted, separately for faces depicted in anterior and profile view, and with and without the inclusion of angular and some distance measurements. Some probe and target faces were correctly matched on both criteria above using the limited sets of measurements. However, some pairs could not be matched using all the measurements in *either* frontal or in profile view. Only when measurements from *both* anterior and profile views were included could the system identify them as a matched pair. From a forensic perspective the number of measurements included in these analyses was higher than would often be obtained in a real case, due mainly to the high quality of the photographs. Furthermore, in many cases, only frontal or profile views would be obtained, not both. These findings indicate that caution should be exercised in the use of photo-anthropometrical analysis, when attempting to 'prove' a match beyond reasonable doubt.

In contrast to the inherent difficulty involved in attempting to confidently state that two different images depict the same person, the use of the *DigitalFace* system reliably distinguished between pairs of faces that had been the subject of the highest number of identification errors in all the experiments reported within the thesis. Indeed, if any reliable differences in facial structure are found between two different images that are not due to variations in photographic equipment or viewpoint; it is possible to conclusively state that they cannot depict the same person.

As was discussed in Chapter 10, one exception would be if a long period of time had elapsed or cosmetic or other changes had altered facial structure. Furthermore, if lower quality images were acquired, as is common in forensic cases, the error associated with landmark identification and distance measurement would have to be integrated within each analysis. To satisfy the recommendations set out by the Court of Appeal in 2003 (*R v Gray*, 2003), it would then be possible to assimilate these values against a database of faces to provide an assessment of whether given the margin of error, the images being compared were of the same or of different people.

**11. 6. Conclusions**

The general public has consistently expressed a positive opinion concerning the intensity of CCTV surveillance in the UK. This is probably based on a belief that it acts as a deterrent to crime, but also that if a crime is committed the offender will be successfully identified and prosecuted. Perhaps counterintuitively, independent evaluations of the effect of CCTV tend to reveal only a moderate reduction in crime rates. Therefore, if CCTV *does* act as a robust deterrent to criminal activity, it is seemingly not possible to demonstrate that effect using current assessment techniques.

Criticism is often directed at CCTV when images of a criminal incident are found to be of poor quality and the appearance of an offender cannot be determined. In these cases, there appears to be a universal belief that if higher quality footage had been obtained, accurate identification of the perpetrator would be certain. However, perhaps again counterintuitively, the experiments reported in this thesis, together with the findings of previous published research, illustrate that the identification of an unfamiliar person depicted even in high-quality images can be error prone. Many offenders will confess to a crime if they are informed by the police that CCTV footage has captured them during the course of their criminal activities. Indeed, many confess if made aware that cameras were surveilling the area, even if the relevant footage has not at the time been acquired. However, these findings have implications in terms of the use of CCTV evidence by the criminal justice system when identity is *disputed*, as a jury may be invited to base a conviction on the resemblance of the offender in video to a defendant (Attorney General's Reference, No. 2 of 2002, 2003). In all cases, jurors will be unfamiliar with the defendant and these findings suggest that if verdicts were based on video evidence, a guilty defendant could escape punishment. Conversely, it is also apparent that it would be difficult for an innocent suspect to prove that they were not depicted in video footage. In addition, an innocent suspect may actually ask the police to inspect CCTV footage from a different locality in an attempt to prove that they were not at the scene of a crime. Again, the results from these experiments demonstrate the difficulty of identity matching in these circumstances.

To prevent miscarriages of justice, judges in summing up would normally warn juries as to the problems involved in identification even when they can directly inspect images. However, the jury may attribute any such warning as a form of 'legalese', a legal necessity in a court room, but to be disregarded as irrelevant when they can 'see for themselves' the resemblance. As the results from all of these studies consistently indicate that human perceptual ability in this unfamiliar face discriminating task can be flawed, other methods of identity verification are necessary and should be pursued.

However, identification testimony from CCTV images may also be heard from people claiming familiarity to the defendant. Most research has demonstrated that a high level of familiarity with a target is associated with high levels of identification accuracy (Bruce et al., 2001; A.M. Burton et al., 1999). Indeed, police forces across the world routinely issue CCTV images to the media in the hope that someone will recognise an offender. However, errors in identification have occurred with even highly familiar people and it is also unclear how much contact is required before a positive identification should be given high credence, although there is an established principle that recognition through prior social interaction is more reliable (*R v Grimer*, 1982).

Testimony may also be accepted from previously unfamiliar witnesses who from extensive frame-to-frame viewing of the images claim to have familiarised themselves to those depicted. The facial learning experiment conducted in this thesis and reported in Chapter 9 was designed to replicate some of these procedures. This revealed that participants asked to closely attend to the video footage over a number of hours, knowing they were to be tested on their recognition of targets, were no better than those who viewed the same footage for a few minutes. This would suggest that there is no advantage in the presentation of this type of testimony. However, each individual legal case will involve different circumstances and to replicate the individual procedures would require specifically designed experiments. Ideally a judge should inform a jury that any such testimony must be *subjective* in nature and can only reflect the personal *opinion* of the witness. Furthermore, regardless of the specific methodology used to inspect the images outside the court, such a witness may be no more able to determine identity from the images than an individual juror within the court.

Finally, due to the difficulty in establishing the identity of an individual in photographic images, it is likely that requests for practitioners to use their specialist expertise in cases of disputed identification will increase, as the number of CCTV cameras continues to rise. The analyses conducted in Chapter 10 demonstrated that with the aid of a new software package, the application of photo-anthropometry could in many cases successfully provide an indication of whether two images depicted the same person. Indeed, in most cases the results were more reliable than human observers. However, it was not possible to reliably individuate some pairs of images without the inclusion of a greater number of facial measurements than would often be available from CCTV footage. In contrast, pairs of *different* faces that had been the subject of misidentifications by many human observers in experiments reported within the thesis were successfully disambiguated by the analyses in most cases. The application of the technique could not guarantee that two different people would not be mistaken for one another, especially if the number of potential measurements was limited, for instance if features were disguised, viewpoint was restricted or image quality was poor. However, it would be advisable for a compliant, genuinely innocent suspect to request that this technique be applied to their case, as it would be more likely to distinguish them from the real culprit, than humans viewing the same footage.

In conclusion, the experiments conducted in this thesis suggest that photo-anthropometrical analysis can provide the most reliable estimation of the likelihood that two different images depict the same person or not. Indeed, it has been demonstrated that the application of this technique is more capable than any human observer of potentially preventing a miscarriage of justice. Many participants in these experiments have been shown to be surprisingly inept at a task that in general most previously considered him or her self to be highly accomplished. It is therefore apparent that caution should be taken in court if the prosecution evidence is based solely on the forensic identification of CCTV images of unfamiliar people.

## 12. 0: List of references

ACPO, (Association of Chief Police Officers of England, Wales and Northern Ireland) (2003). National working practises in facial imagery. Retrieved 15 February 2006 from http://www.acpo.police.uk/asp/policies/data/garvin_facial_imaging_guidelines.doc

Andreasen, N.C., O'Leary, D.S., Arndt, S., Cizadlo, T., Hurtig, R., Rezai, K., Watkins, G.L., Ponto, L.B., & Hichwa, R.D. (1996). Neural substrates of face recognition. Journal of Neuropsychiatry and Clinical Neurosciences, 8, 139 – 146.

Aldridge, J. (1989). The Rotakin — a test target for CCTV security systems: Publication No. 16/89. Police Scientific Development Branch. London: Home Office.

Aldridge, J. (1994). CCTV operations requirement manual version 3.0: Publication number 17/94. Who will be the first to test your CCTV security or safety system? Police Scientific Development Branch. London: Home Office.

Attorney General's Reference, No.2 of 2002 (2003). EWCA Crim 2373

Bannister, J., Fyfe, N.R., & Kearns, A. (1998). CCTV and the city. In C. Norris, J. Moran, and G. Armstrong (Eds.), Surveillance, Closed Circuit Television and Social Control (pp.21 -39). Aldershot: Ashgate.

BBC Health (ND). What's your BMI? Retrieved 7 January 2005 from, http://www.bbc.co.uk/cgi-bin/education/betsie/parser.pl

BBC News (16 August 2003). Tourist mistaken for missing doctor. Retrieved 18 August 2003 from, http://news.bbc.co.uk/1/hi/uk_news/england/merseyside/3156819.stm

BBC News (7 April 2005). CCTV, the latest must-have for the home. Retrieved 15 April 2005 from, http://news.bbc.co.uk/go/pr/fr/-/2/hi/uk_news/magazine/4416587.stm

BBC News. (20 July 2005). Time line: London bombings developments. Retrieved 20 July 2005, from http://news.bbc.co.uk/2/hi/uk_news/4694069.stm.

BBC News (10 October 2005). Trial expert witnesses reviewed. Retrieved 31 October 2005 from, http://news.bbc.co.uk/go/pr/fr/-/1/hi/uk/4317926.stm

Blaney, R.L., & Winograd, E. (1978). Developmental differences in children's recognition memory for faces. Developmental Psychology, 14, 441 – 442.

Bonner, L., & Burton, A.M. (2004). 7-11-year-old children show an advantage for matching and recognizing the internal features of familiar faces: Evidence against a developmental shift. Quarterly Journal of Experimental Psychology, 57A (6), 1019 – 1029.

Bonner, L., Burton, A.M., & Bruce, V. (2003). Getting to know you: How we learn new faces. Visual Cognition, 10 (5), 527 – 536.

Bonner, L., Burton, A.M., Jenkins, R., & McNeil, A. (2003). Meet the Simpsons: Top-down effects in face learning. Perception, 32, 1159 – 1168.

Bornstein, B.H. (1999). The ecological validity of jury simulations: Is the jury still out? Law and Human Behavior, 23 (1), 75 – 91.

Bothwell, R.K., Deffenbacher, K.A., & Brigham, J.C. (1987). Correlation of eyewitness accuracy and confidence. Optimality hypothesis revisited. Journal of Applied Psychology, 72, 691 – 695.

Brace, N.A., Hole, G.J., Kemp, R.I., Pike, G.E., Van Duuren, M., & Norgate, L. (2001). Developmental changes in the effect of inversion: Using a picture book to investigate face recognition. Perception, 30, 85 – 94.

Bramble, S., Compton, D., & Klasen, L. (2001). Forensic image analysis: Version 2. Paper presented at the 13th INTERPOL Forensic Science Symposium, Lyon, France, October 2001.

Brandon, G. (2003). The Police use of CCTV. Paper presented at the 'Is CCTV Working'? Conference, Leicester, UK 18 June 2003.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. Journal of Experimental Psychology: Applied, 8 (1), 44 – 56.

Bray, R.M., & Kerr, N.L. (1979). Use of the simulation method in the study of juror behavior. Law and Human Behavior, 3 (1/2), 107 – 119.

Bray, R.M., & Noble, A.M. (1978). Authoritanarianism and decisions of mock juries: Evidence of jury bias and group polarization. Journal of Personality and Social Psychology, 36, 1424 – 1430.

Brigham, J.C., & Bothwell, R.K. (1983). The ability of prospective jurors to estimate the accuracy of eyewitness identifications. Law and Human Behavior, 7, 19 – 30.

Brigham, J.C., & Wolfskeil, M.P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identification. Law and Human Behavior, 7 (4), 337 – 349.

Bromby, M. (2004). To be taken at face value? Computerised identification. Joseph Bell Centre for Forensic Statistics and Legal Reasoning. Research Report, No. JBC-GCAL-LAW-RR-0004. Retrieved 24 June 2005 from http://www.josephbell.org/

Brown, B. (1995). CCTV in town centres: Three case studies. Police Research Group Crime Detection and Prevention Series, Paper No. 68. London: HMSO.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and circumstances of encounter. Journal of Applied Psychology, 62 (3), 311 – 318.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. British Journal of Psychology, 73, 105 – 116.

Bruce, V., Burton, A.M., Doyle, T., & Dench, N. (1989). Further experiments in the perception of growth in three directions. Perception and Psychophysics, 46, 528 – 536.

Bruce, V., Campbell, R.N., Doherty-Sneddon, G., Import, A., Langton, S., McAuley, S., & Wright, R. (2000). Testing face processing skills in children. British Journal of Developmental Psychology, 18, 319 – 333.

Bruce, V., Henderson, Z., Greenwood, K., Hancock, P.J.B., Burton, AM., & Miller, P. (1999). Verification of face identities from images captured on video. Journal of Experimental Psychology: Applied, 5 (4), 339 - 360.

Bruce, V., Henderson, Z., Newman, C., & Burton, AM. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. Journal of Experimental Psychology: Applied, 7 (3), 207 - 218.

Bruce, V., & Valentine, T. (1988). When a nod's as good as a wink: The role of dynamic information in facial recognition. In M.M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), Practical aspects of memory: Current research and issues: Volume 1 (pp. 169 – 174). Chichester: Wiley.

Bruce, V., Valentine, T., & Baddeley, A. (1987). The basis of the ¾ view advantage in face recognition. Applied Cognitive Psychology, 1, 109 – 120.

Bruce, V., & Young, A.W. (1986). Understanding face recognition. British Journal of Psychology, 77, 305 – 327.

Brunelli, R. & Poggio, T. (1993). Face recognition: Features versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15 (10), 1042 – 1052.

Bull, D. R. (2003). The potential of future image communications technology in crime prevention. In M. Gill (Ed.), CCTV (pp. 141 - 159). Leicester: Perpetuity Press.

Burton, A.M., Bruce, V., & Dench, N. (1993). What's the difference between men and women: Evidence from facial measurement. Perception, 22 (2), 153 – 176.

Burton, A.M., Bruce, V., & Hancock, P.J.B. (1999). From pixels to people: A model of familiar face recognition. Cognitive Science, 23 (1), 1 – 31.

Burton, A. M., Miller, P., Bruce, V., Hancock, P.J.B., & Henderson, Z. (2001). Human and automatic face recognition: A comparison across image formats. Vision Research, 41, 3185 - 3195.

Burton, A. M., Wilson, S., Cowan, M., & Bruce, V. (1999). Face recognition in poor-quality video: Evidence from security surveillance. Psychological Science, 10 (3), 243 - 248.

Burton, P. (2006). Viper – update and current issues. Paper presented at the Nuffield Workshop on Identification Evidence, London, 24th February, 2006.

Campbell, R., Coleman, M., Walker, J., Benson, P.J., Wallace, S., Michelotti, J., & Baron-Cohen, S. (1999). When does the inner-face advantage in familiar face recognition arise and why? Visual Cognition, 6 (2), 197 – 216.

Carey, S., Diamond, R., & Woods, B. (1980). Development of face recognition – a maturational component? Developmental Psychology, 16 (4), 257 – 269.

Catterick, T. (1992). Facial measurements as an aid to recognition. Forensic Science International, 56, 23 – 27.

Chance, J.E., Goldstein, A.G., & Anderson, B. (1986). Recognition memory for infant faces: An analog of the other-race effect. Bulletin of the Psychonomic Society, 24, 257 – 260.

Chiroro, P., & Valentine, T. (1995). An investigation into the contact hypothesis of the own-race bias in face recognition. Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48A, 879 – 894.

Christie, F., & Bruce, V. (1998). The role of dynamic information in the recognition of unfamiliar faces. Memory and Cognition, 26 (4), 780 - 790.

Chung, M. S., & Thomson, D. M. (1995). Development of face recognition. British Journal of Psychology, 86, 55 – 87.

*Church v HMA* (1996) SLT 383

Churchill Insurance Press Release (14 May, 2005). Brits are cheating speed cameras by swapping points with their partners. Retrieved 3 August 2006 from, http://www.churchill.com/pressreleases/14052005.htm

Clutterbuck, R., & Johnston, R.A. (2002). Exploring levels of face familiarity by using an indirect face- matching measure. Perception, 31 (8), 985 - 994.

Clutterbuck, R., & Johnston, R.A. (2004). Demonstrating the acquired familiarity of faces by using a gender-decision task. Perception, 33, 159 – 168.

Clutterbuck, R., & Johnston, R.A. (2005). Demonstrating how unfamiliar faces become familiar using a face matching task. European Journal of Cognitive Psychology, 17 (1), 97 – 116.

Craik, F.I.M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior, 11, 671 – 684.

Crown Prosecution Service. Visual identification of suspects. Retrieved 25 July 2005 from, http://www.cps.gov.uk/legal/section13/chapter_a.html

Cutler, B.L., Fisher, R.P., & Chicvara, C.L. (1989). Eyewitness identification from live versus videotaped lineups. Forensic Reports, 2, 93 – 106.

Cutler, B.L., & Fisher, R.P. (1990). Live lineups, videotaped lineups, and photoarrays. Forensic Reports, 3, 439 – 448.

Cutler, B.L., Penrod, S.D., & Dexter, H.R. (1989). The eyewitness, the expert psychologist, and the law. Law and Human Behavior, 13 (3), 311 – 332.

Cutler, B.L., Penrod, S.D., & Dexter, H.R. (1990). Juror sensitivity to eyewitness identification evidence. Law and Human Behavior, 14 (2), 185 – 191.

Cutler, B.L., Penrod, S.D., & Martens, T.K. (1987). Improving the reliability of eyewitness identification: Putting context into context. Journal of Applied Psychology, 72 (4), 629 – 637.

Cutler, B.L., Penrod, S.D., & Stuve, T.E. (1988). Juror decision making in eyewitness identification cases. Law and Human Behavior, 12 (1), 41 – 55.

Cutting, J.E., & Kozlowski, L.T. (1977). Recognising friends by their walk: Gait perception without familiarity cues. Bulletin of the Psychonomic Society, 9 (5), 353 – 356.

Daily Telegraph (19 July, 2005). Police handed extension to question bomb suspect. Retrieved 19 July 2005 from, http://www.telegraph.co.uk/news/main.jhtml?xml=?news/2005/07/19/utime.xml&sSheet=/portal/2005/07/19/ixportaltop.html

Data Protection Commissioner (2000). CCTV code of practice. London: The Office of the Data Protection Commissioner. Retrieved 20 July 2005 from, http://www.dataprotection.gov.uk/dpr/dpdoc.nsf

*Daubert v. Merrell Dow Pharmaceuticals, Inc* (1993). 509 U.S. 579, 113 S.Ct. 2795

Davies, G.M. (1993). Children's memory for other people: An integrative approach. In C. Nelson (Ed.), Memory and Affect in Development: The Minnesota Symposium on Child Development (pp. 123 – 158. Hillsdale NJ: Erlbaum.

Davies, G.M. (1996a). Mistaken identity: Where law meets psychology head on. Howard Journal of Criminal Justice, 35 (3), 232 – 241.

Davies, G.M. (1996b). Children's identification evidence. In S.L. Sporer, R.S. Malpass and G. Kohnken (Eds.), Psychological Issues in Eyewitness Identification. (pp. 233 – 258). Hillsdale NJ: Erlbaum.

Davies, G., & Thasen, S. (2000). Closed-circuit television: How effective an identification aid? British Journal of Psychology, 91, 411 – 426.

Davies, S. (1999). Spanners in the works: How the privacy movement is adapting to the challenge of Big Brother. In C. Bennett, and R. Grant (Eds.), Visions of Privacy: Policy Choices for the Digital Age. Toronto: University of Toronto Press.

Davis, J.H., Kerr, N.L., Atkin, R.S., Holt, R., & Meek, D. (1975). The decision processes of 6- and 12-person mock juries assigned unanimous and two-thirds majority rules. Journal of Personality and Social Psychology, 32 (1), 1 – 14.

Davis, J.H., Stasson, M., Ono, K., & Zimmerman, S. (1988). Effects of straw polls on group decision making: Sequential voting pattern, timing and local majorities. Journal of Personality and Social Psychology, 55, 918 – 926.

Dent, H.R. (1977). Stress as a factor influencing person recognition in identification parades. Bulletin of the British Psychological Society, 30, 339 – 340.

Dent, H.R., & Stephenson, S. (1979). Identification evidence: Experimental investigations of factors affecting the reliability of juvenile and adult witnesses. In D. Farrington, K. Hawkins, and L. Lloyd-Bostock (Eds.), Psychology, Law and Legal Processes (pp. 195 – 206). Atlantic Highlands, NJ: Humanities Press.

Devlin, Lord P. (1976). Report to the Secretary of State for the Home Department on the Departmental Committee on Evidence of Identification in Criminal Cases. London: HMSO.

Diamond, R., & Carey, S. (1977). Developmental changes in the representation of faces. Journal of Experimental Child Psychology, 23, 1 – 22.

Diamond, R., Carey, S., & Back, K.J. (1983). Genetic influences on the development of spatial skills during early adolescence. Cognition, 13, 167 – 185.

Diamond, S.S., & Zeisel, H. (1974). A courtroom experiment on juror selection and decision-making. Personality and Social Psychology Bulletin, 1, 276 – 277.

Dillehay, R.C., & Nietzel, M.T. (1985). Juror experience and jury verdicts. Law and Human Behavior, 9 (2), 179 – 191.

Ditton, J. (2000). Crime and the city: Public attitudes towards open-street CCTV in Glasgow. British Journal of Criminology, 40, 692 - 709.

Ditton, J., & Short, E. (1998). Evaluating Scotland's first town centre CCTV scheme. In C. Norris, J. Moran, and G. Armstrong (Eds.), Surveillance, Closed Circuit Television and Social Control (pp. 155 - 174.). Aldershot: Ashgate.

Ditton, J., Short, E., Phillips, S., Norris, C., & Armstrong, G. (1999). The effect of closed circuit television cameras on recorded crime rates and public concern about crime in Glasgow. The Scottish Office Central Research Unit 65(14).

DR Nyheder Online (15 July 2005). Police: Surveillance will help terror research. Retrieved 18 July 2005 from, http://www.dr.dk/nyheder/fremmedsprog/English/article.hhtml?articleID=264793

Dror, I.E., Charlton, D., & Peron, A.E. (2006). Contextual information renders experts vulnerable to making erroneous identifications. Forensic Science International, 156, 74 – 78.

DVLA (Driver and Vehicle Licensing Authority, 2004). Annual report 2003 – 2004. Retrieved 4 August 2005 from, http://dvla.gov.uk/public/annual_report_04/ar_ch2.htm

Ebbesen, E.B., & Bowers, R.J. (1974). Proportion of risky to conservative arguments in a group discussion and choice shifts. Journal of Personality and Social Psychology, 29, 316 – 327.

Egan, D., Pittner, M., & Goldstein, A.G. (1977). Eyewitness identification: Photographs vs. live models. Law and Human Behavior, 1 (2), 199 - 206.

Ellis, H.D. (1992). The development of face processing skills. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 335 (1273), 105 – 111.

Ellis, H.D., Shepherd, J.W., & Davies, G.M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. Perception, 8, 431 – 439.

European Commission (2005). Biometrics at the Frontiers: Assessing the impact on society. For the European Parliament Commission on Citizens' Freedoms and Rights, Justice & Home Affairs (LIBE). European Commission, Joint Research Centre: Institute for Prospective Technological Studies. http://www.jrc.es.htm

Farrington, D.P., & Painter, K.A. (2003). How to evaluate the impact of CCTV on crime. In M. Gill (Ed.), CCTV (pp. 67 - 79). Leicester: Perpetuity Press.

Feraud, R., Bernier, O.J., Viallet, J.E., & Collobert, M. (2001). A fast and accurate face detector based on neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23 (1), 42 – 53.

Fitzpatrick, M. (2002, 14 October 2002). Better than text? Independent, 14 October 2002, p. 10.

Flin, R. H. (1980). Age effects in children's memory for unfamiliar faces. Developmental Psychology, 16 (4), 373 – 374.

Flin, R. H. (1985a). Development of face recognition: An encoding switch? British Journal of Psychology, 1, 123 – 134.

Flin, R.H. (1985b) Development of visual memory: An early adolescent regression. Journal of Early Adolescence, 5, 259 – 266.

Flight, S., van Heerwaarden, Y., & van Soomeren, P. (2003). Does CCTV displace crime? An evaluation of the evidence and a case study from Amsterdam. In M. Gill (Ed.), CCTV (pp. 93 - 107). Leicester: Perpetuity Press.

Foot, H., & Sanford, A. (2004). The use and abuse of student participants. The Psychologist, (17), 5, 256 – 259.

Forensic Science Service (2004) Annual report and accounts 2003 - 2004. Retrieved 1 August 2005 from, http://www.forensic.gov.uk/forensic_t/inside/about/annual.htm

Foucault, M. (1977). Discipline and Punish: The Birth of the Prison. Allen Lane: London.

Fulton, A., & Bartlett, J.C. (1991). Young and old faces in young and old heads: The factor of age in face recognition. Psychology and Ageing, 6 (4), 623 – 630.

Fyfe, N.R., & Bannister, J. (1996). City watching: Closed circuit television surveillance in public spaces. Area, 28 (1), 37 – 46.

Gallagher, C. (2004). CCTV and human rights: The fish and the bicycle? An examination of Peck v. United Kingdom (2003) 36 E.H.R.R. 41. http://www.surveillance-and-society.org/cctv.htm

George, P.A., & Hole, G.J. (1995). Factors influencing the accuracy of age estimates of unfamiliar faces. Perception, 24, 1059 – 1073.

Gill, M., Allen, J., Bryan, J., Kara, D., Little, R., Waples, S., Spriggs, A., Argomaniz, J., Jessiman, P., Kilworth, J., & Swain, D. (2005). The impact of CCTV: Fourteen case studies. Home Office, Research Development and Statistics Directorate. Online Report 15/05. London: Home Office. http://www.homeoffice.gov.uk/rds

Gill, M., & Loveday, K. (2003a). What do offenders think about CCTV? In M. Gill (Ed.), CCTV (pp. 81 - 91). Leicester: Perpetuity Press.

Gill, M., & Loveday, K. (2003b). What do offenders think of CCTV? Paper presented at the 'Is CCTV Working?' Conference, Leicester, UK 18 June 2003.

Gill, M., Smith, P., Spriggs, A., Argomaniz, J., Allen, J., Follett, M., Jessiman, P., Kara, D., Little, R., & Swain, D. (2003). National evaluation of CCTV: Early findings on scheme implementation – effective practice guide. Home Office Research, Development and Statistics Directorate: Development and Practice Report, 7. London: Home Office.

Goldstein, A.G., Chance, J.E., & Schneller, G.R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. Bulletin of the Psychonomic Society, 27, 71 – 74.

Goodman, G.S., & Reed, R.S. (1986). Age differences in eyewitness testimony. Law and Human Behavior, 10 (4), 317 – 332.

Goold, B.J. (2003). Public area surveillance and police work: The impact of CCTV on police behaviour and autonomy. Surveillance & Society, 1(2), 191 - 203. http://www.surveillance-and-society.org/cctv.htm

Graham, S. (1998). Towards the fifth utility? On the extension and normalisation of public CCTV. In C. Norris, J. Moran, and G. Armstrong (Eds.) Surveillance, Closed Circuit Television and Social Control (pp. 89 – 112). Aldershot: Ashgate.

Gras, M.L. (2004). The legal regulation of CCTV in Europe. Surveillance & Society, 2 (2/3), 216 - 229. http://www.surveillance-and-society.org/cctv.htm

Griffiths, M. (2003). Town centre CCTV: An examination of crime reduction in Gillingham, Kent. London: Home Office. Retrieved 24 Jun 2005 from http://www.crimereduction.gov.uk/cctv.33.htm

Grimm, F. (28 February 2006). Big brother shows his good side in court. The Miami Herald. Retrieved 28 February 2006 from http://www.miami.com/mld/miamiherald/13977381.htm

Groscup, J.L., Penrod, S.D., Studebaker, C.A., Huss, M.T., & O'Neill, K.M. (2002). The effects of Daubert on the admissibility of expert testimony in state and federal cases. Psychology, Public Policy and Law, 8 (4), 339 – 372.

Harper, B., & Latto, R. (2001). Cyclopean vision, size estimation, and presence in orthostereoscopic images. Presence, 10 (3), 312 – 330.

Harries, R. (1999). The Cost of Criminal Justice. Home Office Development and Statistics Directorate, Research Findings No. 103.

Hastie, R., Penrod, S.D., & Pennington, N. (1983). Inside the Jury. Cambridge, MA: Harvard University Press.

Heisele, B., Ho, P., Wu, J., & Poggio, T. (2003). Face recognition: Component-based versus global approaches. Computer Vision and Image Understanding, 91 (1 - 2), 6 - 21.

Helten F., & Fischer, B. (2004). Urban Eye, Working Paper No. 13. What do people think about CCTV? Findings from a Berlin survey. http://www.urbaneye.net/results/results.htm

Henderson, Z., Bruce, V., & Burton, A.M. (2001). Matching the faces of robbers captured on video. Applied Cognitive Psychology, 15, 445 – 464.

Henn, M., & Weinstein, M. (2002). Do you remember the first time? Paper presented at the PSA Conference, Aberdeen, 5 – 7 April, 2002. Retrieved 10 August 2005 from, http://gsr.ntu.ac.uk/docs/PSA%20paper%202002%20_Apr%204_%203.pdf

*Herbecq v Belgium* (1998) App. No. 32200/96, decision of 14th January 1998

Hill, H., & Bruce, V. (1996). Effects of lighting on the perception of facial surfaces. Journal of Experimental Psychology: Human Perception & Performance, 22 (4), 986 – 1004.

Hill, H., & Pollick, F.E. (2000). Exaggerating temporal differences enhances recognition of individuals from point light displays. Psychological Science, 11 (3), 223 – 228.

Hill, H., Schyns, P.G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. Cognition, 62 (2), 201 – 222.

Hjelmas, E., & Low, B.K. (2001). Face detection: A survey. Computer Vision and Image Understanding, 83 (3), 236 – 274.

Hockley, W.E., Hemsworth, D.H., & Consoli, A. (1999). Shades of the mirror effect: Recognition of faces with and without sunglasses. Memory & Cognition, 27 (1), 128 – 138.

Honore, C. (12 April 2004). Britain's Big Brother. Houston Chronicle. Retrieved 15 April 2004 from, http://www.chron.com/cs/CDA/printstory.mpl/front/2499854

House of Lords: Science and Technology Committee; 5th Report (1997/1998). Digital images as evidence. Retrieved 1 August 2005 from, http://www.publications.partliament.uk/pa/ld200102/ldinfo/reports.htm#8199798

House of Lords: Science and Technology Committee; 8th Report (1997/1998). Digital images as evidence: Government response. Retrieved 1 August 2005 from, http://www.publications.partliament.uk/pa/Id200102/Idinfo/reports.httm#8199798

Hogan, J. (2003). Smart software linked to CCTV can spot dubious behaviour. New Scientist, 11 July 2003.

Honess, T., & Charman, B. (1992). Closed Circuit Television in Public Places. In Crime Prevention Unit Paper: Number 38. London: Police Research Group.

Hopkins, N., & Hall, S. (2000). Fostering hate that turned quiet son into a murderer. The Guardian: 1 July, 2000. Retrieved 2 October 2003 from, http:www.guardian.co.uk/bombs/Story/0,2763,338618,00.html

Hosch, H.M., Beck, E.L., & McIntyre, P. (1980). Influence of expert testimony regarding eyewitness accuracy on jury decisions. Law and Human Behavior, 4 (4), 287 – 296.

Hurley, D.J., Nixon, M.S., & Carter, J.N. (2005). Force field feature extraction for ear biometrics. Computer Vision and Image Understanding, 98, 491 – 512.

Independent (23 August 2005). Police rely on 6,000 cameras across Tube network to cut crime. Retrieved 23 August 2005 from, http://news.independent.co.uk/uk/transport/article307679.ece

Independent (10 January 2006). Blair vs. Yobs. The Independent, 10 January 2006.

Introna, L.D. & Wood, D. (2004). Picturing algorithmic surveillance: The politics of facial recognition systems. Surveillance & Society, 2 (2/3), 177- 198. http://www.surveillance-and-society.org.cctv.htm

Irving, R. (8 August 2005). Obsolete CCTV units 'put London in danger'. The Times. Retrieved 9 August 2005 from, http://business.timesonline.co.uk/article/0,,8209-1725923,00.html

Iscan, M.Y. (1993). Introduction to techniques for photographic comparison: Potentials and problems, in M.Y. Iscan and R.P. Helmer (Eds.), Forensic Analysis of the Skull (pp. 57 – 70). New York: Wiley-Liss.

Isenberg, D.J. (1986). Group polarization: A critical review and meta-analysis. Journal of Personality and Social Psychology, 50 (6), 1141 – 1151.

Kalven, H., & Zeisel, H. (1966). The American Jury. Chicago: University of Chicago Press.

Kaplan, M.F., & Miller, C.E. (1977). Judgements and group discussion: Effects of presentation and memory factors on polarization. Sociometry, 40, 337 – 343.

Kathimerini (26 August 2005). Minister irate at CCTV ruling. Retrieved 26 August 2005 from, http://www.ekathimerini.com/4dcgi/_w_articles_politics_100010_26/08/2005_60045

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces – the role of shape from shading and pigmentation cues. Perception, 25, 37 – 52.

Kemp, R., Towell, N., & Pike, G. (1997). When seeing should not be believing: Photographs, credit cards and fraud. Applied Cognitive Psychology, 11, 211 – 222.

Kerr, N.L., Harmon, D.L., & Graves, J.K. (1982). Independence of multiple verdicts by jurors and juries. Journal of Applied Social Psychology, 12, 12 – 29.

Kerr, N.L., MacCoun, R.J., & Kramer, G.P. (1996). Bias in judgment. Comparing individuals and groups. Psychological Review, 103 (4), 687 – 719.

Kerr, N.L., Nerenz, D., & Herrick, D. (1979). Role playing and the study of jury behavior. Sociological Methods and Research, 7, 337 – 355.

Kirby, M., & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1), 103 – 108.

Knight, W. (2006). Activists highjack public CCTV signal. Retrieved 9/01/2006 from http://newscientist.com/article.ns?id=dn8530&print=true

Knight, B., & Johnston, A. (1997). The role of movement in face recognition. Visual Cognition, 4, 265 – 273.

Lander, K., & Chuang, L. (2005). Why are moving faces easier to recognize? Visual Cognition, 12 (3), 429 – 442.

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. Applied Cognitive Psychology, 15 (1), 101 - 116.

Laughery, K.R., Rhodes, B.T., & Batten, G.W. (1981). Computer-guided recognition and retrieval of facial images. In G.M. Davies, J.W. Shepherd and H.D. Ellis (Eds.), Perceiving and Remembering Faces. New York: Academic Press.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition – women's faces make the difference. Brain and Cognition, 50, 121 – 128.

Lichfield, J. (2006). Isabelle faces the world. The Independent, 7 February 2006, p. 18.

Light, L.L., Kayra-Stuart, F., & Hollander, S. (1979). Recognition memory for typical and unusual faces. Journal of Experimental Psychology: Human Learning and Memory, 5, 212 – 228.

Lindsay, R.C.L., Pozzulo, J.D., Craig, W., Lee, K., & Corber, S. (1997). Simultaneous lineups, sequential lineups, and showups: Eye-witness identification decisions of adults and children. Law and Human Behavior, 21 (4), 391 – 404.

Liu, C.H., & Chaudhuri, A. (2003). Face recognition with perspective transformation. Vision Research, 43, 2393 – 2402.

Liu, C.H., Seetzen, H., Burton, A.M., & Chaudhuri, A. (2003). Face recognition is robust with incongruent image resolution: Relationship to security video images. Journal of Experimental Psychology: Applied, 9 (1), 33 - 41.

Lomell, H.M., Saetnan, A.R., & Wiecek, C. (2003). Urban Eye, Working Paper No. 9. Flexible technology, structured practises: Surveillance operations in 14 Norwegian and Danish organisations. http://www.urbaneye.net/results/results.htm

Loula, F., Prasad, S., Harber, K., & Shiffrar, M. (2005). Recognizing people from their movement. Journal of Experimental Psychology: Human Perception and Performance, 31 (1), 210 – 220.

Loveday, K., & Gill, M. (2003). The impact of monitored CCTV in a retail environment: What CCTV operators do and why. In M. Gill (Ed.), CCTV (pp. 109 - 125). Leicester: Perpetuity Press.

Luus, C.A.E., & Wells, G.L. (1991). Eyewitness identification and the selection of distracters for lineups. Law and Human Behavior, 15 (1), 43 – 57.

McCahill, M., & Norris, C. (2003a). Estimating the extent, sophistication and legality of CCTV in London. In M. Gill (Ed.), CCTV (pp. 51 - 65). Leicester: Perpetuity Press.

McCahill, M., & Norris, C. (2003b). Estimating the extent, sophistication and legality of CCTV in London. Paper presented at the Is CCTV Working? Conference, Leicester, UK. 18 June 2003.

McCahill, M, & Norris, C. (2003c). Urban Eye, Working Paper No. 10. CCTV systems in London: Their structures and practises. http://www.urbaneye.net/results/results.htm

McCahill, M, & Norris, C. (2003d). Urban Eye, Working Paper No. 3. CCTV in Britain. http://www.urbaneye.net/results/results.htm

Macaskill, M. (10 April 2005). CCTV-free countryside offers richer pickings for urban criminals. The Times. Retrieved 15 April 2005 from, http://www.timesonline.co.uk/printFriendly/0,,1-1506-1562911-1506,00.html

MacCoun, R.J., & Kerr, N.L. (1988). Asymmetric influence in mock jury deliberation: Jurors' bias for leniency. Journal of Personality and Social Psychology, 54, 21 – 33.

Madsen, D.B. (1978). Issue importance and choice shifts: A persuasive arguments approach. Journal of Personality and Social Psychology, 36, 1118 – 1127.

Maguire, M. (1998). Restraining Big Brother? The regulation of surveillance in England and Wales. In C. Norris, J. Moran, and G. Armstrong, (Eds.), Surveillance, Closed Circuit Television and Social Control (pp.229 - 241). Aldershot: Ashgate.

Malone, D.R., Morris, H.H., Kay, M.C., & Levin, H.S. (1982). Prosopagnosia; a double dissociation between the recognition of familiar and unfamiliar faces. Journal of Neurology, Neurosurgery, and Psychiatry, 45, 820 – 822.

Mann, V.A., Diamond, R., & Carey, S. (1979). Development of voice recognition. Parallels with face recognition. Journal of Experimental Child Psychology, 27, 153 – 165.

Mardia, K.V., Coombs, A., Kirkbride, J., Linney, A., & Bowie, J.L. (1996). On statistical problems with face identification from photographs. Journal of Applied Statistics, 23 (6), 655 – 675.

Mead, L. (1998). Usage of video recordings in surveillance, the value of such as evidence and potential problems which can arise. Paper presented at the 13th Annual BILETA Conference, Dublin, 27 – 28 March 1998. Retrieved 1 August 2005 from http://www.bileta.ac.uk/98papers/mead.html

Meissner, C.A., & Brigham, J.C. (2001). Thirty years of investigating the own-race bias in memory for faces. Psychology, Public Policy and Law, 7 (1), 3 – 35.

Metzger, M.M. (2001). Which transformations of stimuli are the most disruptive to facial recognition? Perceptual and Motor Skills, 92 (2), 517 – 526.

Michelmore, K. (1 December 2005).Murdoch witness rejects 'facial map'. Retrieved 3/12/2005 from http://www.theaustralian.news.com.au/printpage/0,5942,17426074,00.html

Moenssens, A.A. (1999). Identifying individuals by ear photographs or ear prints. Paper presented at the First International Conference on Forensic Human Identification, London, 23 – 26 October 1999.

Murphy, T. (1999). The admissibility of CCTV evidence in criminal proceedings. International Review of Law and Computers, 13 (3), 383 – 404.

Myers, D.G. (1978). Polarising effects of social comparison. Journal of Experimental Social Psychology, 14, 554 – 563.

Myers, D.G., Bruggink, J.B., Kersting, R.C., & Schlosser, B.A. (1980). Does learning others' opinions change one's opinion? Personality and Social Psychology Bulletin, 6, 253 – 260.

Myers, D.G., & Lamm, H. (1976). The group polarization phenomenon. Psychological Bulletin, 83, 602 – 627.

Nagao, D.H., & Davis, J.H. (1980). The effects of prior experience on mock juror case judgments. Social Psychological Quarterly, 43, 190 – 199.

*Neil v Biggers* (1972) 409 U.S. 188

Nemeth, C. (1977). Interactions between jurors as a function of majority vs. unanimity decision rules. Journal of Applied Social Psychology, 7 (1), 38 – 56.

Newcombe, N., & Lie, E. (1995). Overt and covert recognition of faces in children and adults. Psychological Science, 6 (4), 241 – 245.

Nixon, M.S., & Carter, J.N. (2004). Advances in automatic gait recognition. In Proceedings of IEEE Face and Gesture Analysis 2004, FG04, Seoul, South Korea, pp. 11 – 16. Retrieved 5 August 2005 from, http://www.eprints.ecs.soton.ac.uk/10069/

Norris, C., & Armstrong, G. (1998). Introduction: Power and vision. In C. Norris, J. Moran, and G. Armstrong (Eds.), Surveillance, Closed Circuit Television and Social Control (pp. 3 - 18). Aldershot: Ashgate.

Norris, C., & Armstrong, G. (1999). The Maximum Surveillance Society: The rise of CCTV. Oxford: Berg.

Norris, C., McCahill, M., & Wood, D. (2004). Editorial. The growth of CCTV: A global perspective on the international diffusion of video surveillance in publicly accessible space. Surveillance & Society, 2 (2/3), 110 - 135. http://www.surveillance-and-society.org/cctv.htm

*North Yorkshire Trading Standards Department v Williams* (1994) TLR 22.11.94

Novesti (Russian News & Information Agency (18 July, 2005) Moscow metro to be equipped with surveillance cameras. Retrieved 19 July 2005 from, http://en.rian.ru/russia/20050718/40927397.html

Number 10 (31 December, 2004). Prime Minister delivers New Year message. Retrieved 31 December 2004 from, http://www.number-10.gov.uk/output/Page6853.asp

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. Perception, 30 (6), 755 - 764.

O'Rourke, T.E., Penrod, S.D., Cutler, B.L., & Stuve, T.E. (1989). The external validity of eyewitness identification research: Generalising across subject populations. Law and Human Behavior, 13 (4), 385 – 395.

Oxlee, G. (ND). Imagery analysis aspects of recognition and identification. Retrieved from www.kalagate.co.uk/fss_gathering_paper.pdf

Painter, K. A., & Farrington, D. P. (1997). The crime reducing effect of improved street lighting: The Dudley project. In R. V. Clarke, (Ed.), Situational Crime Prevention: Successful Case Studies, 2nd Edition. (pp. 209-226). Guilderland, NY: Harrow and Heston.

Painter, K. A., & Farrington, D. P. (1999). Street lighting and crime: Diffusion of benefits in the Stoke-on-Trent project. In K. A. Painter, & N. Tilley, (Eds.), Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention (pp. 77 - 122). Monsey, N.Y: Criminal Justice Press.

Parker, J. (2000). Total Surveillance. London: Piatkus.

Parker, J.F., & Carranza, L.E. (1989). Eyewitness testimony of children in target-present and target-absent lineups. Law and Human Behavior, 13 (2), 133 – 149.

Parker, J.F., Haverfield, E., & Baker-Thomas, S. (1986). Eyewitness testimony of children. Journal of Applied Social Psychology, 16 (4), 287 – 301.

Parker, J.F., & Ryan, V. (1993). An attempt to reduce guessing behavior in children's and adults' eyewitness identifications. Law and Human Behavior, 17 (1), 11 – 26.

Patterson, K.E., & Baddeley, A.D. (1977). When face recognition fails. Journal of Experimental Psychology: Human Learning and Memory, 3 (4), 406 – 417.

Pawson, R., & Tilley, N. (1994). What works in evaluation research? British Journal of Criminology, 34 (3), 391 – 406.

Peters, D.P. (1991). The influence of stress and arousal on the child witness. In J. Doris (Ed.), The Suggestibility of Children's Recollections (pp. 60 – 76). Washington DC: American Psychological Association.

POST (Parliamentary Office of Science and Technology). (2001). Biometrics and security. Postnote, No. 165. London: HMSO. http://www.parliament.uk/post/home.htm

POST (Parliamentary Office of Science and Technology). (2002). CCTV. Postnote, No. 175. London: HMSO. http://www.parliament.uk/post/home.htm

POST (Parliamentary Office of Science and Technology). (2004). Speed Cameras. Postnote, No. 218. London: HMSO. http://www.parliament.uk/post/home.htm

Peace, K. (1997). Crime Prevention. In M. Maguire, Morgan, R. and Reiner, R. (Ed.), The Oxford Handbook of Criminology, 2nd Edition. (pp. 963 -995). Oxford: Oxford University Press.

*Peck v United Kingdom* (2003) 36 E.H.R.R. 41

Peterborough Evening Telegraph (12 January 2005). Neighbours from hell – on CCTV. Retrieved 17 January 2005 from, http://www.peterboroughtoday.co.uk/ViewArticle2.aspx

Phillips, C. (1999). A review of CCTV evaluations: Crime reduction effects and attitudes towards its use. In K. Painter and N. Tilley (Eds.). Surveillance of Public Space: CCTV, Street Lighting and Crime Prevention, Crime Prevention Studies, Vol. 10 (pp. 123 – 155). Monsey NY: Criminal Justice Press.

Phillips, P. J., Grother, P., Micheals, R.J., Blackburn, D.M., Tabassi, E., & Bone, J.M. (2003). Face recognition vendor test 2002: Overview and Summary. Retrieved: 21 July 2003. http://frvt.or/FRVT2002/documents.htm

Pike, G., Brace, N., & Kynan, S. (2002). The visual identification of suspects: Procedures and practise. Briefing Note No. 2/02. Policing and Reducing Crime Unit, Home Office Research, Development and Statistics Directorate, London: Home Office.

Privacy International (22 July 1997). CCTV frequently asked questions. Retrieved 8 August 2005 from http://www.privacyinternational.org/article.shtml?cmd(347)=x-347-61925&als(theme)

Pramateftakis, M., Oelbaum, T., & Diepold, K. (2004). Authentication of MPEG-4 surveillance video. International Conference on Image Processing, 24-27 Oct. 2004. Volume: I, On page(s): 33-37 Vol. 1 Singapore

*R v Blenkinsop* (1995) 1 Cr.App.R(S)7

*R v Bowden* (1993) Crim LR 379

*R v Caldwell and Dixon* (1993) CLR 862

*R v Campbell* (1996) TLR 20.02.96

*R v Christou* (1992) 1 QB 979

*R v Clare and Peach* (1995) 2 Cr App R 333

*R v Clarke* (1995) 2 Cr App R 425

*R v Dodson and Williams* (1984) 79 Cr App R 220

*R v Fowden and White* (1982) Crim LR 588

*R v Gardner* (2004) EWCA Crim 1639

*R v Gray* (2003) EWCA Crim 1001

*R v Grimer* (1982) Crim LR 674

*R v Hookway* (1999) Crim LR 750

*R v Kerrigan* (1998) Court of Appeal, 11 June 1998

*R v Loveridge and others* (2001) EWCA Crim 973

*R v McNamara* (1996) CLR 750

*R v Pattinson and Exley* (1996) 1 Cr App 51

*R v Stockwell* (1993) 97 Cr App R 260

*R v Thomas* (1994) Crim.L.R. 128, CA

*R v Turnbull and others* (1976) 3 All ER 549

*R v Turner* (1975) 60 Cr App R 80 CA

Ranger, S. (2006). Police build national mugshot database. Retrieved 18/01/2006 from http://www.silicon.com/publicsector/0,3800010409,39155636,00htm

Rattner, A. (1988). Convicted but innocent: Wrongful conviction and the criminal justice system. Law and Human Behavior, 12, 283 – 293.

Reed, J. (17 February, 2006). Super-fast 3D pictures set to be an eye-opener for security industry. Retrieved 20 February 2006, from http://www.yorkshiretoday.co.uk/viewarticle2.aspx?sectionid=55&articleid=1356677

Reeve, A. (1998). The panopticisation of shopping: CCTV and leisure consumption. In C. Norris, J. Moran, and G. Armstrong, (Eds.), Surveillance, Closed Circuit Television and Social Control (pp. 69 - 87.). Aldershot: Ashgate.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. Perception, 17, 43 – 63.

Rhodes, G., Brennan, S., & Carey, S. (1987). Identification and ratings of caricatures: Implications for mental representations of faces. Cognitive Psychology, 19, 473 – 497.

Sanders, G.S., & Baron, R.S. (1977). Is social comparison irrelevant for producing choice shifts? Journal of Experimental Social Psychology, 13, 303 – 314.

Sandys, M., & Dillehay, R.C. (1995). First-ballot votes, predeliberation dispositions, and final verdicts in jury trials. Law and Human Behavior, 19, 175 – 195.

Sarker, D. (12 April 2004). The new face of biometrics. Federal Computer Week. Retrieved 15 April 2004 from, http://fcw.com/

Scheck, B., Neufeld, P., & Dwyer, J. (2000). Actual innocence: Five days to execution and other dispatches from the wrongly convicted. New York: Doubleday.

Schiff, W., Banka, L., & de Bordes Galdi, G.D. (1986) Recognising people seen in events via dynamic "mug shots". American Journal of Psychology, 99 (2), 219 – 231.

Schofield, D., & Goodwin, L. (2004). Error in 3D capture of facial images. Talk given at ICARIS, London, 29 Oct 2004.

Scientific Working Group on Imaging Technology (SWGIT) (2005). Best practices for forensic image analysis. Forensic Science Communications, 7 (4), 1 – 12.

Searcy, J.H., Bartlett, J.C. & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. Memory & Cognition, 27, 538 – 552.

Shapiro, P.N., & Penrod, S. (1986). Meta-analysis of facial identification studies. Psychological Bulletin, 100, 139 - 156.

Shepherd, J.W., Ellis, H.D., & Davies, G.M. (1982). Identification evidence: A psychological evaluation. Aberdeen: University of Aberdeen Press.

Silicon.com (29 March 2004). London Underground looks at 'smart' anti-terror monitoring. Retrieved 5 April 2004 from, http://www.silicon.com/software/security/0,39024655,39119621,00.htm

Simmons, J., & Dodd, T. (2003). Crime in England and Wales: 2002/2003. London: HMSO.

Simons, D.J., & Ambinder, M.S. (2005). Change blindness: Theory and consequences. Current Directions in Psychological Science, 14 (1), 44 – 48.

Simons, D.J. and Levin, D.T. (1997) Change blindness. Trends in Cognitive Science, 1, 261–267

Sinha, P. (1996). Symmetry sensing through computer vision and a facial recognition system. Forensic Science International, 77, 27 – 36

Sinha, P. (1998). A symmetry perceiving adaptive neural network and facial image recognition. Forensic Science International, 98 (1 – 2), 67 – 89.

Sivarajasingam V., Shepherd, J.P., & Matthews, K. (2003). Effect of urban closed circuit television on assault injury and violence detection. Injury Prevention, 9 (4), 312 - 316.

Skinns, D. (1997). Annual Report of the Safety in Doncaster Evaluation: October 1995 - September 1996). Doncaster: Safety in Doncaster.

Skinns, D. (1998). Crime reduction, diffusion and displacement: Evaluating the effectiveness of CCTV. In C. Norris, J. Moran, and G. Armstrong (Eds.), Surveillance, Closed Circuit Television and Social Control (pp. 175 - 188.). Aldershot: Ashgate.

*Smith v The Queen* (2001). HCA 50.

Smith, A.D., & Winograd, E. (1978). Adult age differences in remembering faces. Developmental Psychology, 14, 443 – 444.

Snodgrass, J.G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. Journal of Experimental Psychology, General, 117, 34 – 50.

Socolinsky, D.A., Selinger, A., & Neuheisel, J.D. (2003). Face recognition with visible and thermal infrared imagery. Computer Vision and Image Understanding, 91 (1 – 2), 72 – 114.

Soppe, H.J.G. (1986). Children's recognition of unfamiliar faces: Developments and determinants. International Journal of Behavioral Development, 9, 219 – 233.

Sporer, S.L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. Psychological Bulletin, 118 (3), 315 – 327.

Squires, P., & Measor, L. (1996). Closed Circuit TV Surveillance and Crime Prevention in Brighton: Half Yearly Report. Brighton: Health and Social Policy Research Centre, University of Brighton.

Stevenage, S.V., Nixon, M.S., & Vince, K. (1999). Visual analysis of gait as a cue to identity. Applied Cognitive Psychology, 13, 513 – 526.

Steyn, M. (2005). Big Brother caused our hooded youth. Daily Telegraph, 17 May 2005. Retrieved 19 May 2005 from, http://www.opinion.telegraph.co.uk

Sutton, A., & Wilson. D. (2004). Open-street CCTV in Australia: The politics of resistance and expansion. Surveillance & Society, 2 (2/3), 310 - 322. http://www.surveillance-and-society.org/cctv.htm

Takacs, B. (1998). Comparing face images using the modified Hausdorff distance. Pattern Recognition, 31 (12), 1873 – 1881.

Taylor v The Chief Constable of Cheshire (1987) 84 CR.APP.R. 191

Taylor, N. (2002). State surveillance and the right to privacy. Surveillance & Society, 1 (1), 66 – 85. http://www.surveillance-and-society.org/cctv.htm

Technical Working group for Eyewitness Evidence (1999). Eyewitness evidence: A guide for law enforcement. Washington, DC: United States Department of Justice, Office of Justice Programs.

Terry, R.L. (1993). How wearing eyeglasses affects facial recognition. Current Psychology, 12 (2), 151 – 162.

Terry, R.L. (1994). Effects of facial transformations on accuracy of recognition. Journal of Social Psychology, 134 (4), 483 – 492.

The Times (15 July, 2005). Speed camera U-turn as 500 sites rejected. Retrieved 3 August 2006 from http://www.timesonline.co.uk/article/0,,2-1694858,00.html

The Times (16 December, 2005). Speed camera benefits overrated. Retrieved 3 August 2006 from http://www.timesonline.co.uk/article/0,,2-1934085,00.html

The Times (24 May, 2006). New speed camera puts more drivers in the frame. Retrieved 3 August 2006 from http://www.timesonline.co.uk/article/0,,2-2194624,00.html

Treleven, E. (25 February, 2006). Caught on camera: Surveillance devices make it tough for crooks. Retrieved 28 February 2006 from http://www.madison.com/

Troscianko, T., Holmes, A., Stillman, J., Mirmehdi, M., Wright, D., & Wilson, A. (2004). What happens next? The predictability of natural behaviour viewed through CCTV cameras. Perception, 33, 87 – 101.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. Journal of Cognitive Neuroscience, 3 (1), 71 – 86.

UK Passport Service (2005). UK Passport Service: Improving passport security and tackling ID fraud. Retrieved 1 August 2005 from, http://www.ukpa.gov.uk/press_240305.asp

USA Today (17 July, 2005). Cities opening more video surveillance eyes. USA Today. Retrieved 19 July 2005 from, http://www.usatoday.com/news/nation/2005-07-17-cameras-cities_x.htm

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. British Journal of Psychology, 79, 471 – 492.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion and race in face recognition. Quarterly Journal of Experimental Psychology, 43A (2), 161 – 204.

Valentine, T., & Bruce, V. (1986a). Recognising familiar faces: The role of distinctiveness and familiarity. Canadian Journal of Psychology, 40, 300 – 305.

Valentine, T., & Bruce, V. (1986b). The effects of distinctiveness in recognising and classifying faces. Perception, 15, 525 – 535.

Valentine, T., & Endo (1992). Towards an exemplar model of face processing: The effects of race and distinctiveness. Quarterly Journal of Experimental Psychology, 44A, 671 – 703.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. Applied Cognitive Psychology, 17, 969 – 993.

Vanezis, P., Lu, D., Cockburn, J., Gonzalez. A., McCombe, G., Trujillo, O., & Vanezis, M. (1996). Morphological classification of facial features in adult Caucasian males based on an assessment of photographs. Journal of Forensic Science, 41 (5), 786 – 791.

Verdant Technologies (2004). Hi-lighter: Digital image event recorder (press release). Retrieved 5 October 2004 from, http://verdant-tech.co.uk

Vinokur, A., & Burnstein, E. (1978). Novel argumentation and attitude change: The case of polarization following group discussion. European

Want, S.C., Pascalis, O., Coleman, M., & Blades, M. (2003). Recognising people from the inner or outer parts of their faces: Developmental data concerning 'unfamiliar' faces. British Journal of Developmental Psychology, 21, 125 – 135.

Washington Post. (8 October 2005). Security camera new witness. Retrieved 10 October 2005 from,
http://www.washingtonpost.com/wp-yn/content/article2005/10/07/AR2005100701895

Webster, W.R. (2004). The diffusion, regulation and governance of closed-circuit television in the UK. Surveillance & Society, 2 (2/3), 230 - 250. http://www.surveillance-and-society.org/cctv.htm

Wells, G.L. (1993). What do we know about eyewitness identification? American Psychologist, 48, 553 – 571.

Wells, G.L., Malpass, G.L., Lindsay, R.C., Fisher, R.P., Turtle, J.W., & Fulero, S.M. (2000). From the lab to the police station: A successful application of eyewitness research. American Psychologist, 55 (6), 581 – 598.

Wells, G.L., & Windschitl, P.D. (1999). Stimulus sampling and social psychological experimentation. Personality and Social Psychology Bulletin, 25 (9), 1115 – 1125.

Welsh, B.C., & Farrington, D.P. (2002). Crime prevention effects of closed circuit television: A systematic review. Home Office Research Study No. 252. London: Home Office.

Wiecek, C., & Saetnan, A.R. (2002). Working Paper No. 5. Geographies of visibility: Zooming in on video surveillance systems in Oslo and Copenhagen. http://www.urbaneye.net.htm

Wikipedia (3 August, 2006). Road-rule enforcement camera. Retrieved 3 August 2006 from http://en.wikipedia.org/wiki/speed-camera

Williams, C.A. (2003). Police surveillance and the emergence of CCTV in the 1960's. In M. Gill (Ed.), CCTV (pp. 9 - 21). Leicester: Perpetuity Press.

Wilson, D.W., & Donnerstein, E. (1977). Guilty or not guilty? A look at the "simulated" jury paradigm. Journal of Applied Social Psychology, 7, 175 – 190.

Winge, S., & Knutsson, J. (2003). An evaluation of the CCTV scheme at Oslo Central Railway station. In M. Gill (Ed.), CCTV (pp. 127- 140). Leicester: Perpetuity Press.

Wise, R.A., & Safer, M.A. (2004). What US judges know and believe about eyewitness testimony. Applied Cognitive Psychology, 18, 427 – 443.

Wiskott, L., Fellous, J.M., Kruger, N., & von der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. IEEE Transactions on Image Analysis and Machine Intelligence, 19 (7), 775 – 779.

Woffinden, B. (2004). Earprint landed innocent man in jail for murder. Retrieved 25 January 2004 from, http://www.guardian.co.uk/uk_news/story/0,3604,1129414,00.html

Wogalter, M.S., & Laughery, K.R. (1987). Face recognition: Effects of study to test maintenance and change of photographic mode and pose. Applied Cognitive Psychology, 1 (4), 241 – 253.

Wright, D.B., & McDaid, A.T. (1996). Comparing system and estimator variables using data from real lineups. Applied Cognitive Psychology, 10, 75 – 84.

Wright, D.B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. Acta Psychologica, 114, 101 – 114.

Wright, D.B., & Stroud, J.N. (2002). Age differences in lineup identification accuracy: People are better with their own age. Law and Human Behavior, 26, (6), 641 – 654.

Yoshino, M., Noguchi, K., Atsuchi, M., Kubota, S., Imaizumi, K., Thomas, C.D.L., & Clement, J.G. (2002). Individual identification of disguised faces by morphometrical matching. Forensic Science International, 127 (1 - 2), 97 – 103.

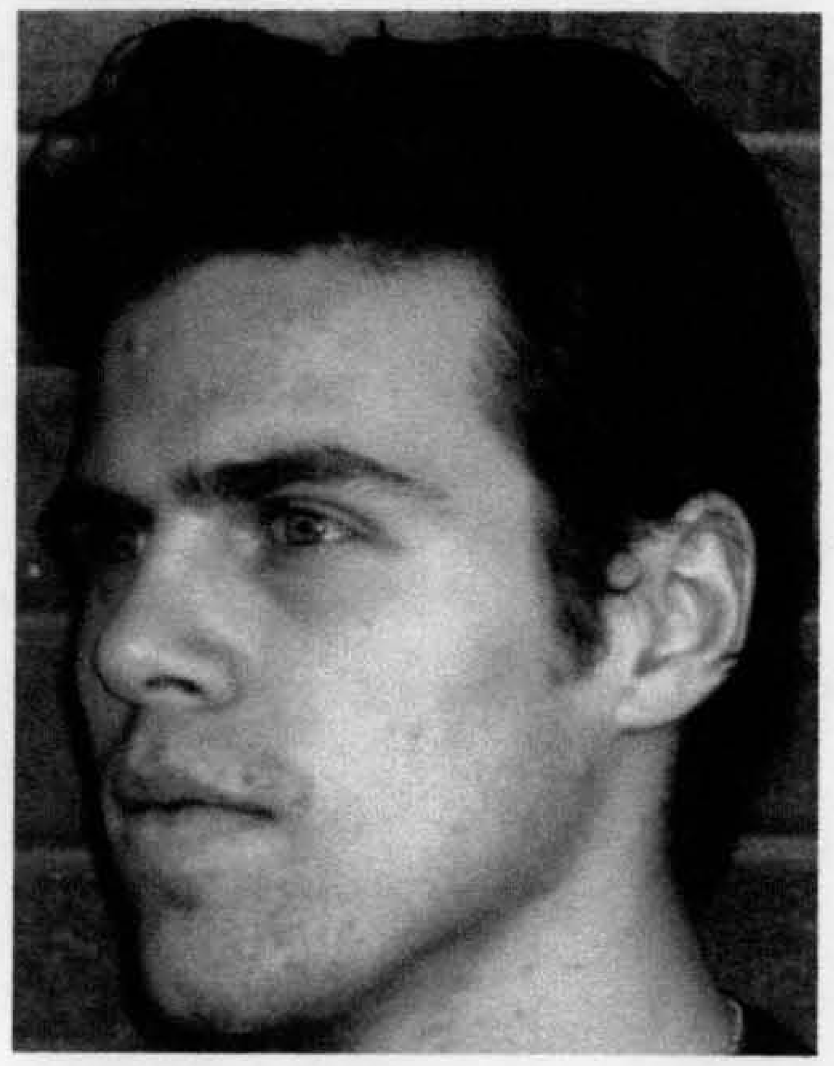
Yoshino, M., Matsuda, H., Kubota, S., Imaizumi, K., & Miyasaka, S. (2000). Computer-assisted facial image identification system using a 3-D physiognomic range finder. Forensic Science International, 109 (3), 225 – 237.

Young, A.W., Hay, D.C., & Ellis, A.W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognising people. British Journal of Psychology, 76, 495 – 523.

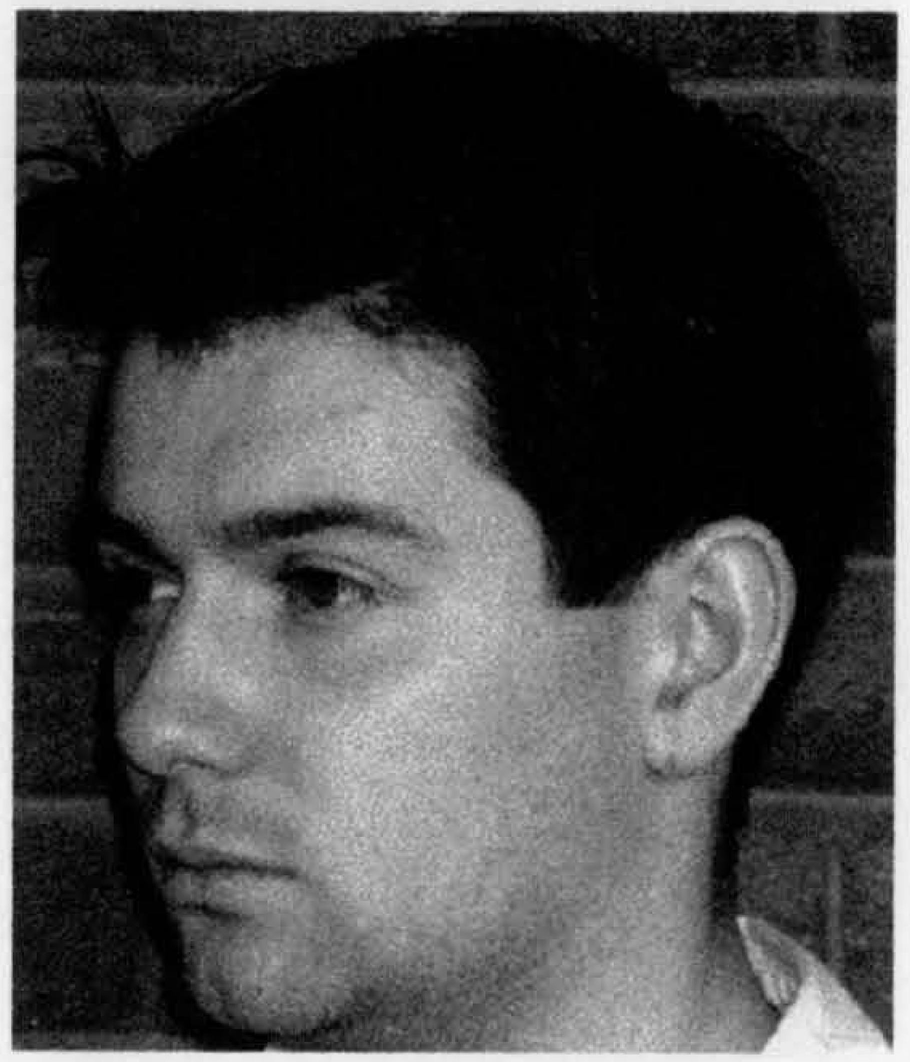
Young, A.W., Hay, D.C., McWeeny, K.H., Flude, B.M., & Ellis, A.W. (1985). Matching familiar and unfamiliar faces on internal and external features. Perception, 14 (6), 737 - 746.

Young, A.W., Newcombe, F., De Haan, E.H.F., Small, M., & Hay, D.C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. Brain, 116, 941 – 959.

## Appendix A.1: Three-quarters facial portraits used in Experiment 4.1 (sub-group 1)

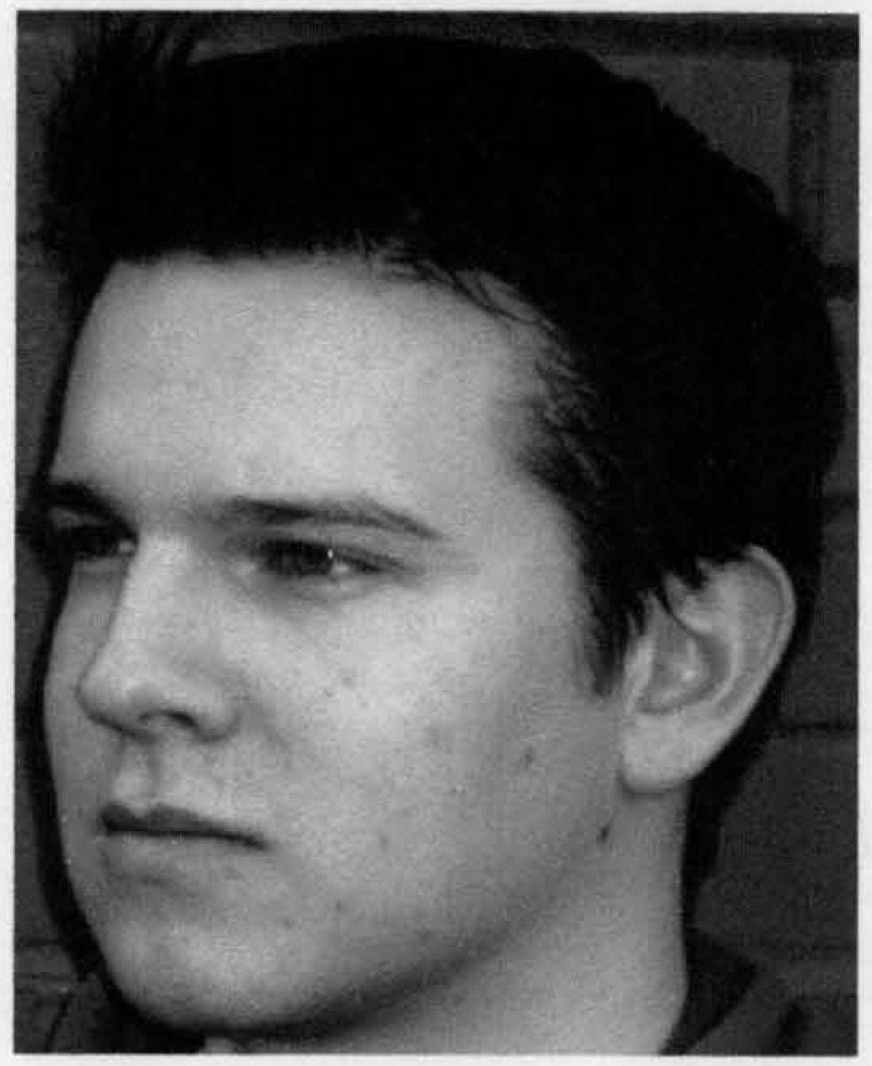


*Actor 01*

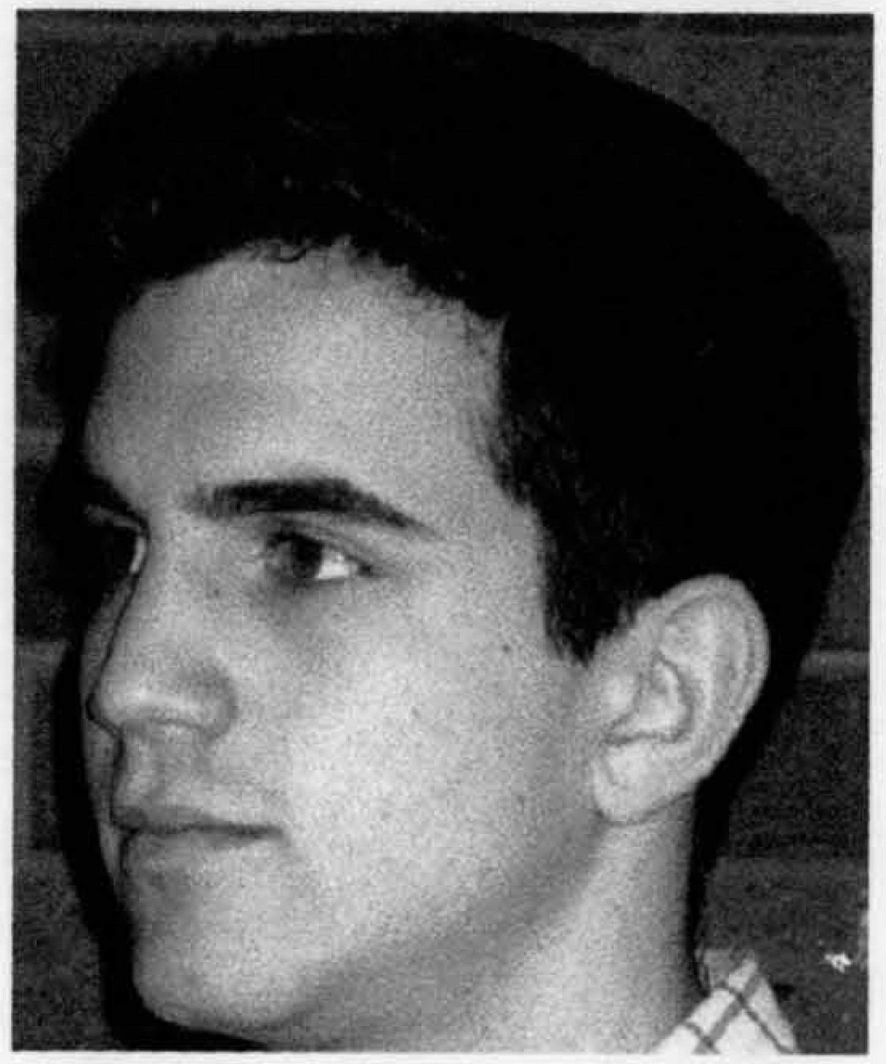


*Actor 03*

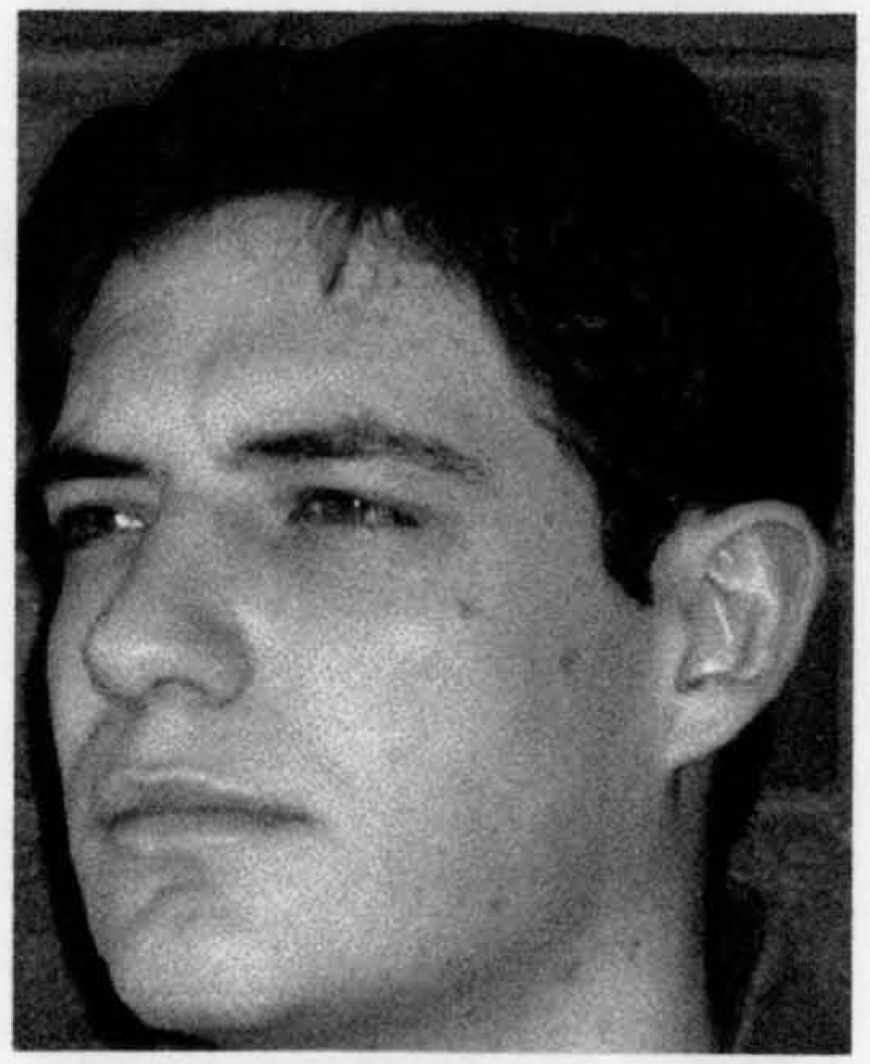
*Actor 05*



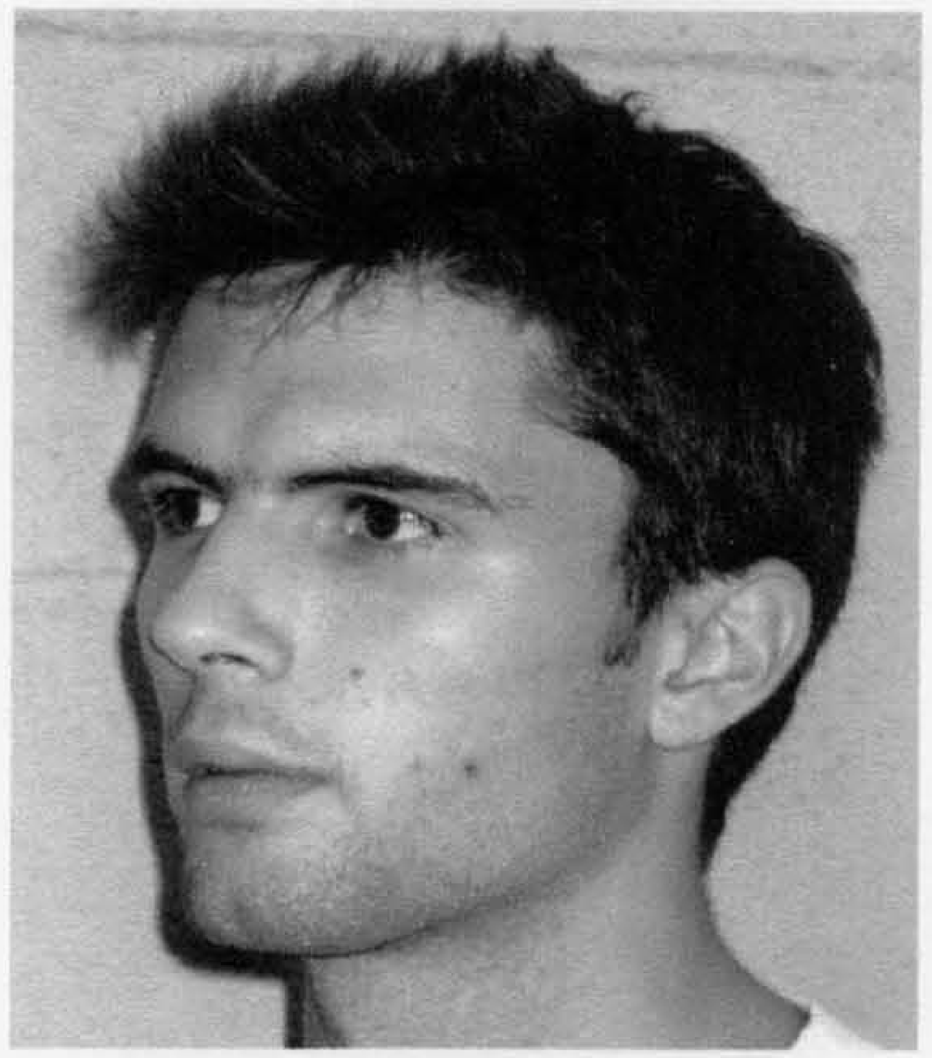
*Actor 07*



*Actor 02*



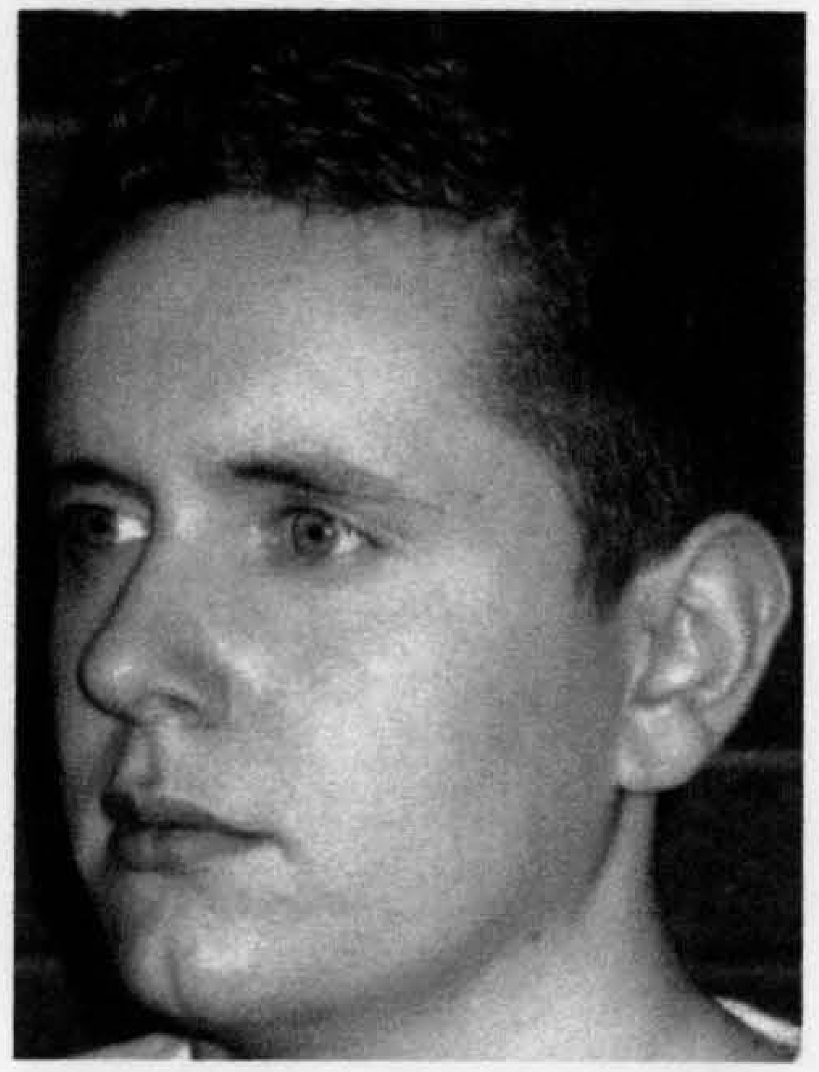
*Actor 04*



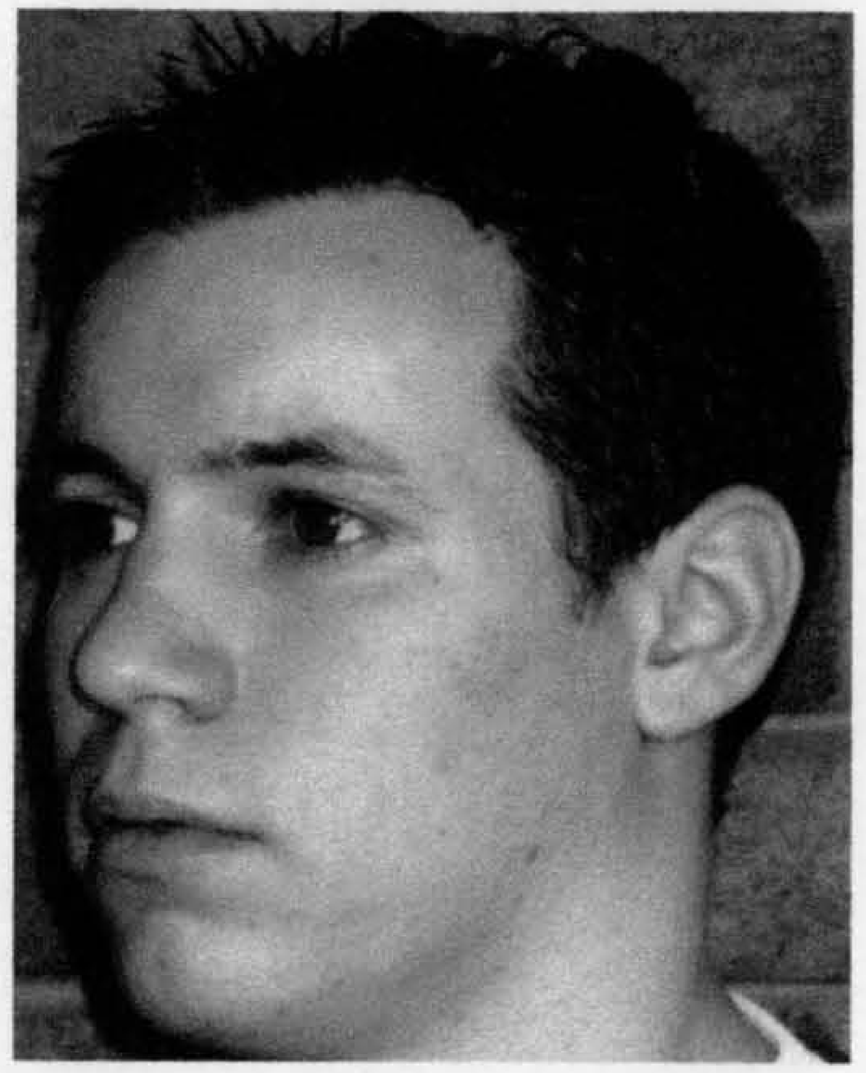
*Actor 06*

*Figure A1: Photographs used in sub-group array number 1 in Experiment 4.1. Actors 01, 02, 03 & 04 were depicted in video in target present conditions. Actor 07 replaced the target for target absent conditions. See section 4.1.1.2 for details.*

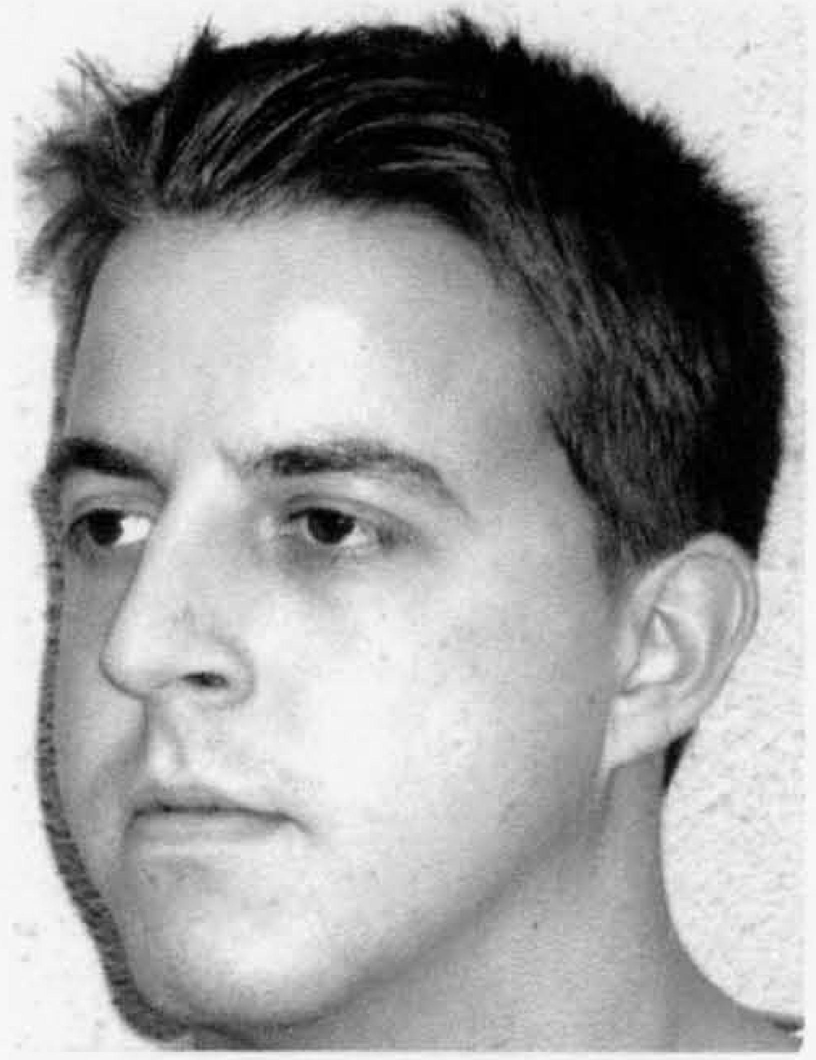
## Appendix A.2: Three-quarters facial portraits used in Experiment 4.1 (sub-group 2)



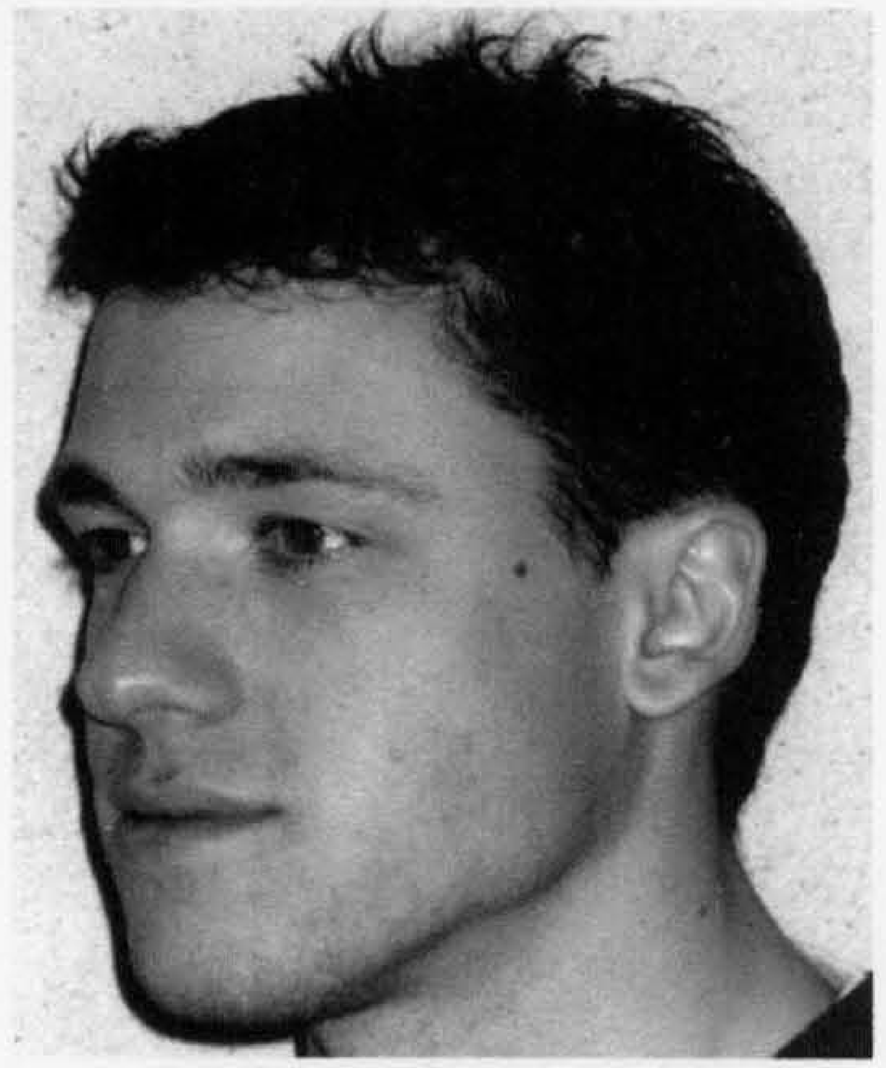
*Actor 08*



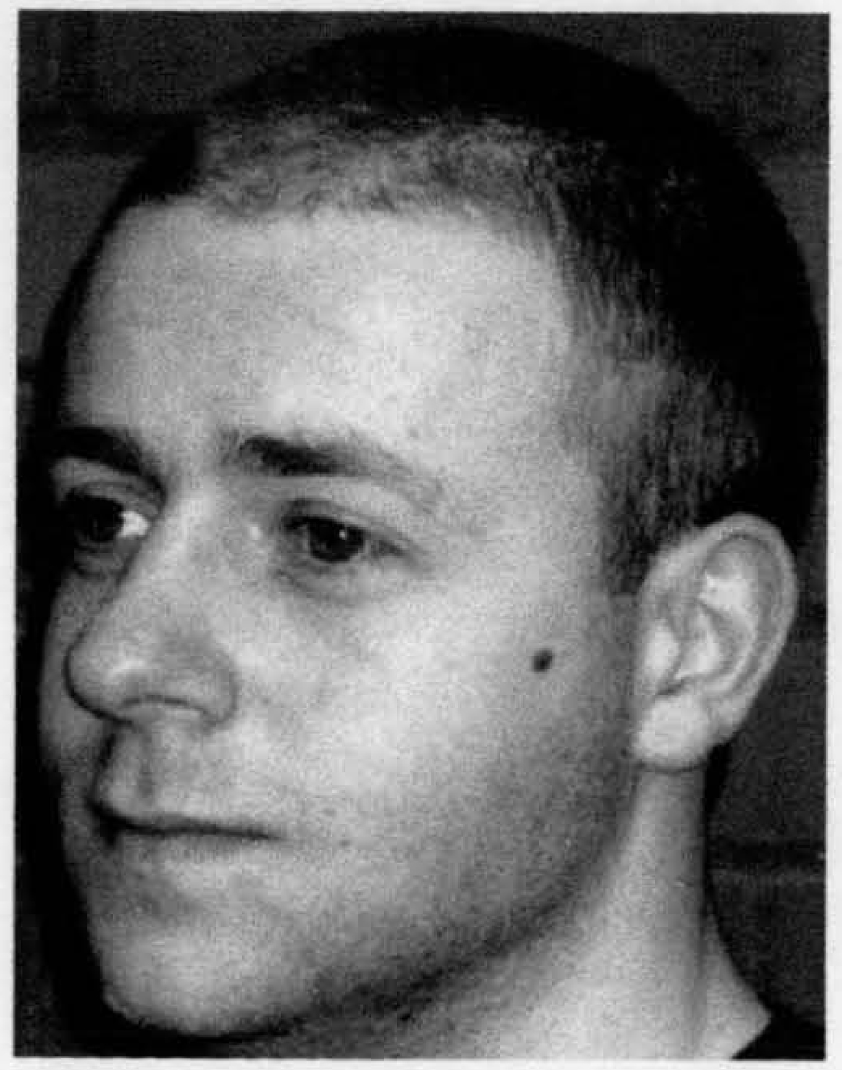
*Actor 10*



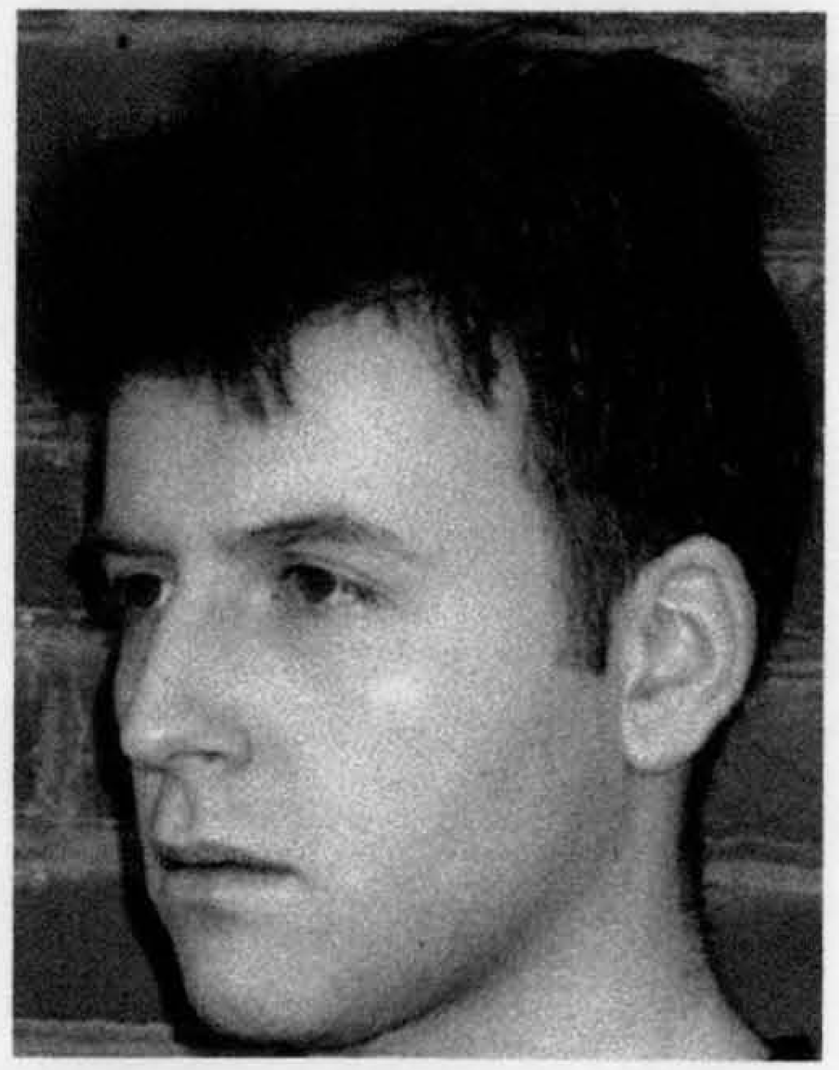
*Actor 12*



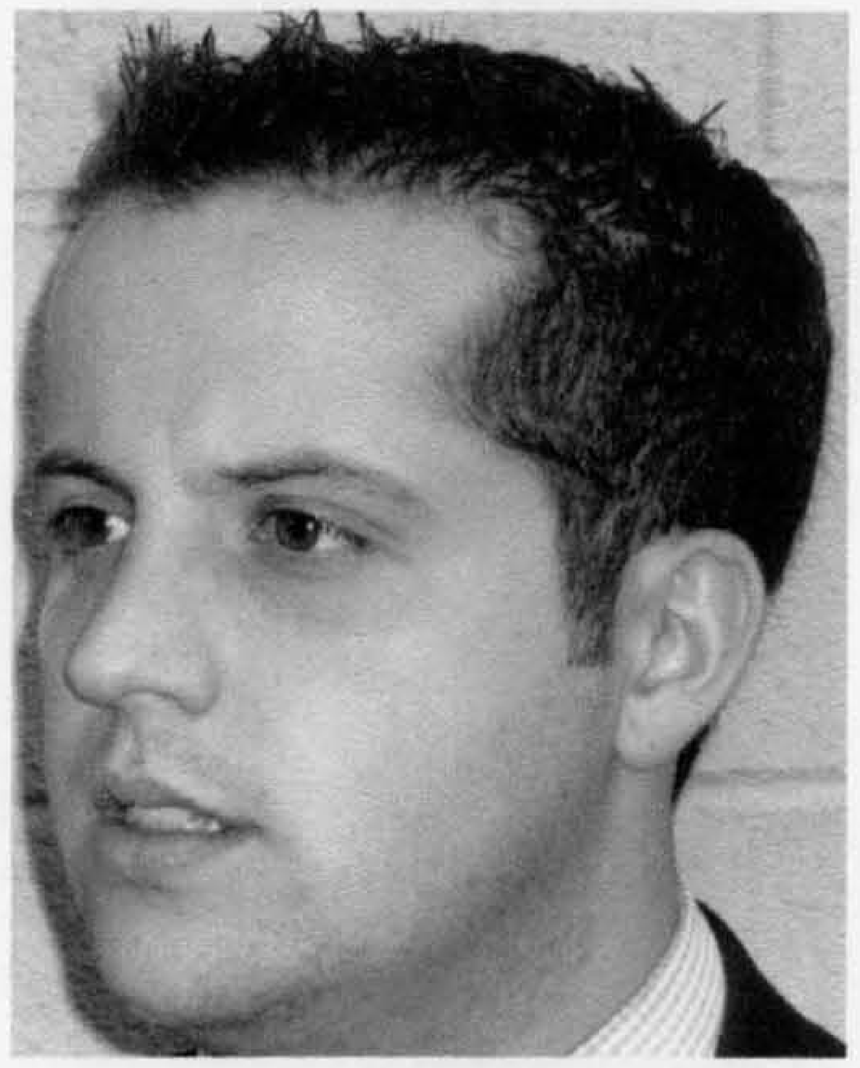
*Actor 14*



*Actor 09*



*Actor 11*

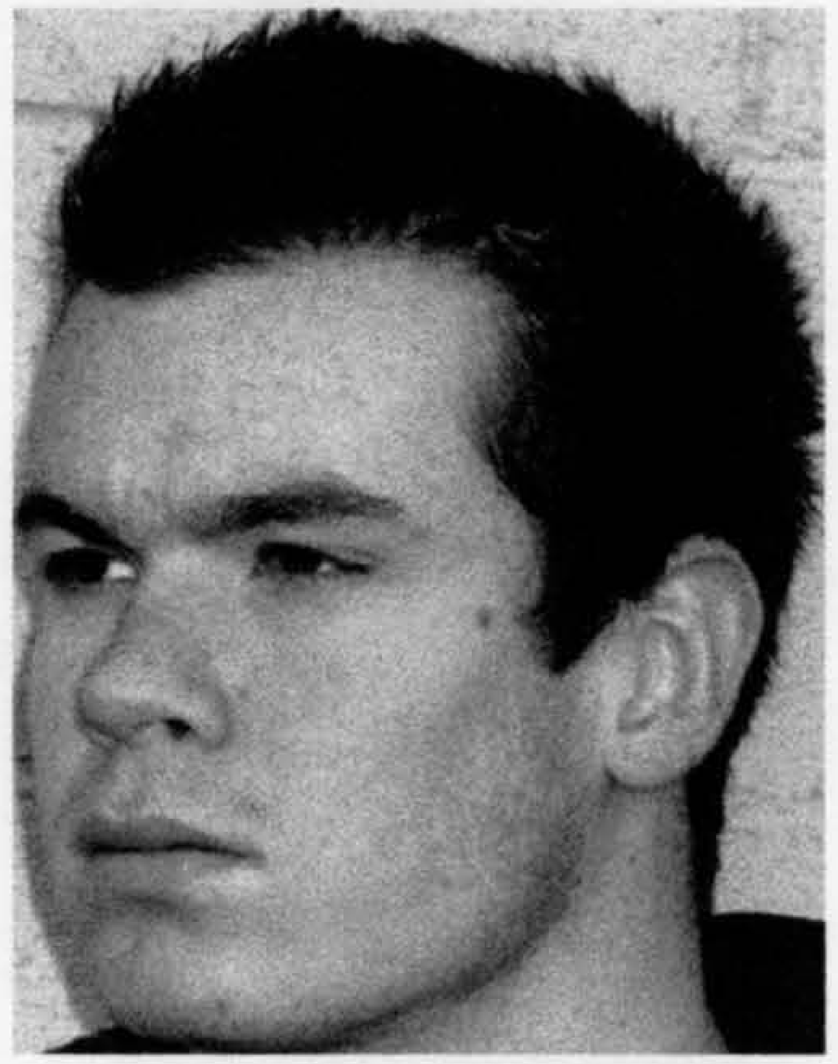


*Actor 13*

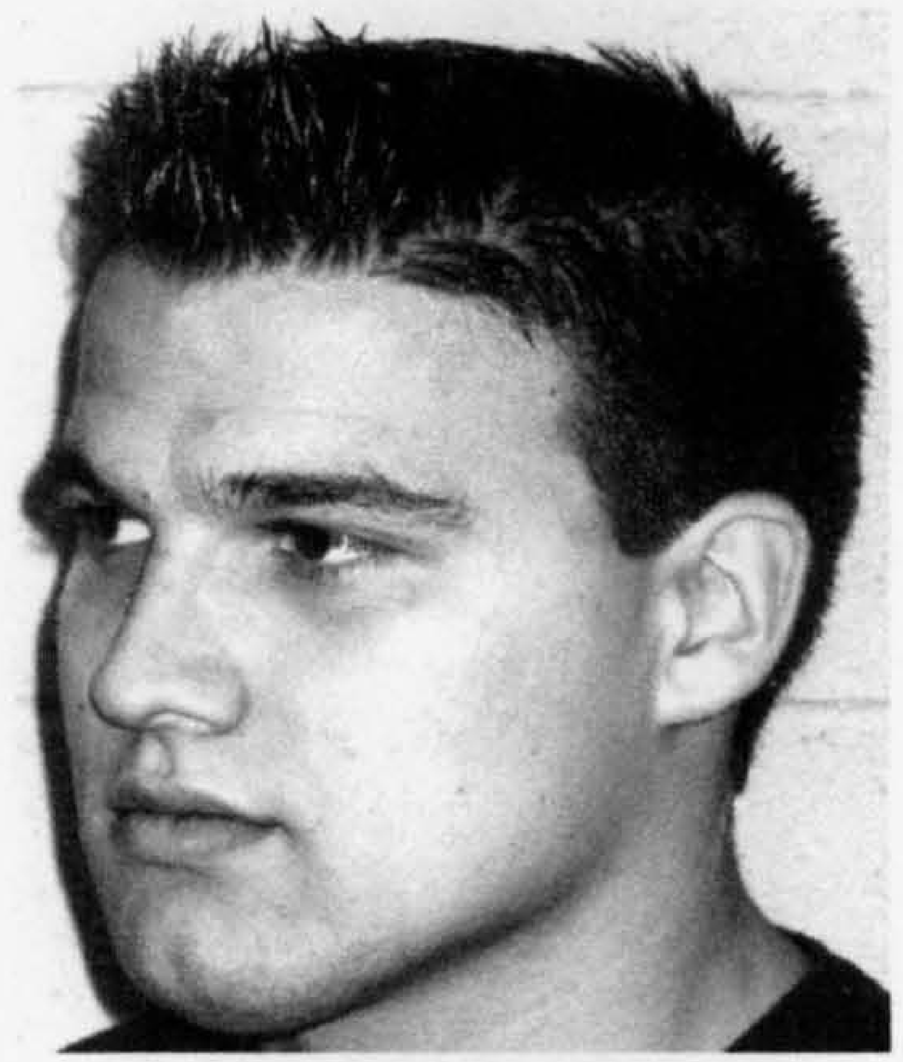
*Figure A2: Photographs used in sub-group array number 2 in Experiment 4.1. Actors 08, 09, 10 & 11 were depicted in video in target present conditions. Actor 14 replaced the target for target absent conditions. See section 4.1.1.2 for details.*

**Appendix A.3**: **Three-quarters facial portraits used in Experiment 4.1 (sub-group 3)**

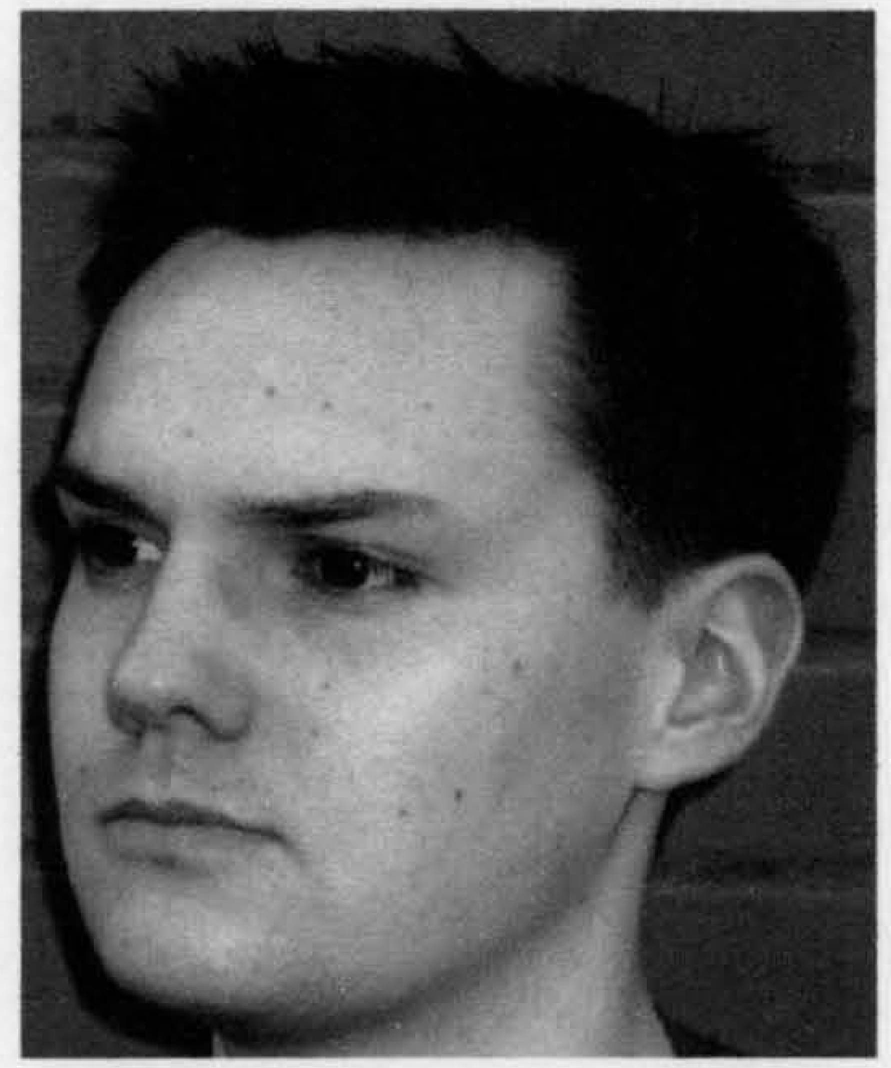
*Actor 15*



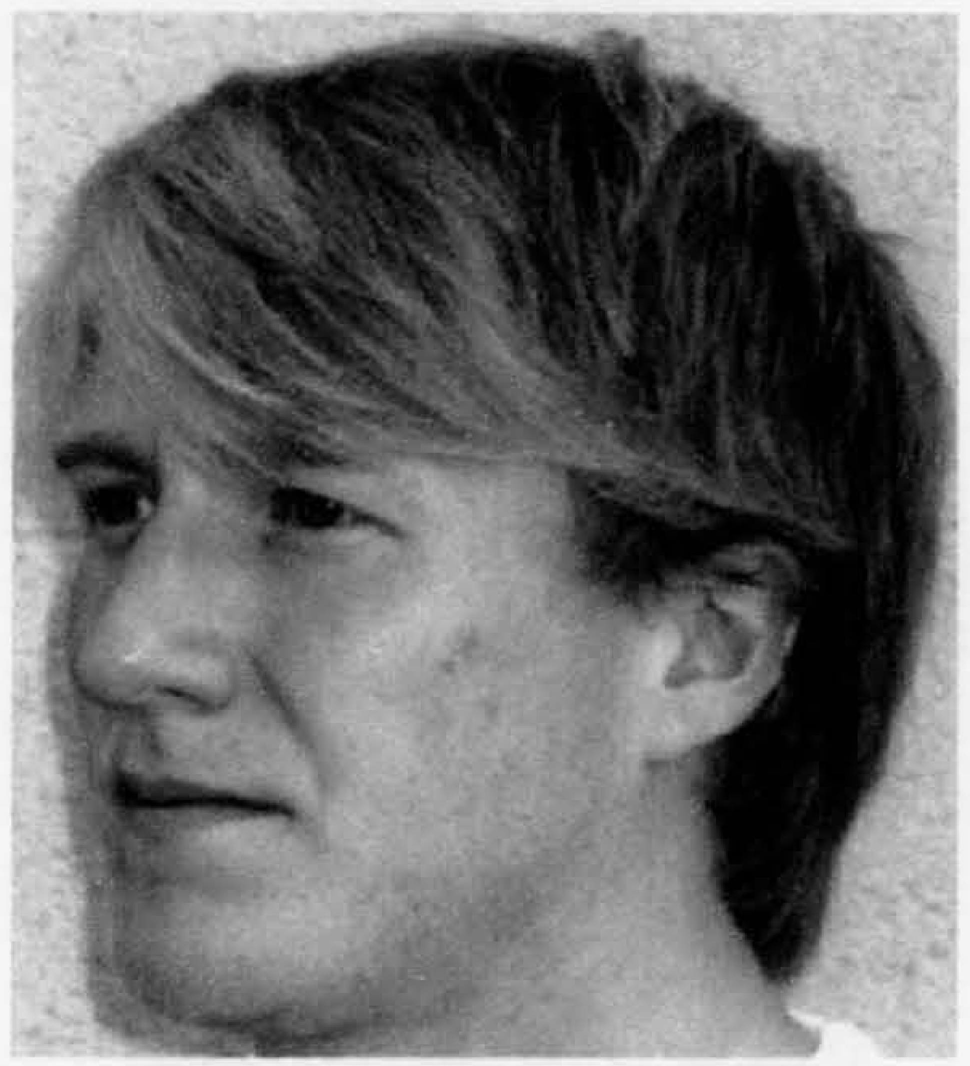
*Actor 17*



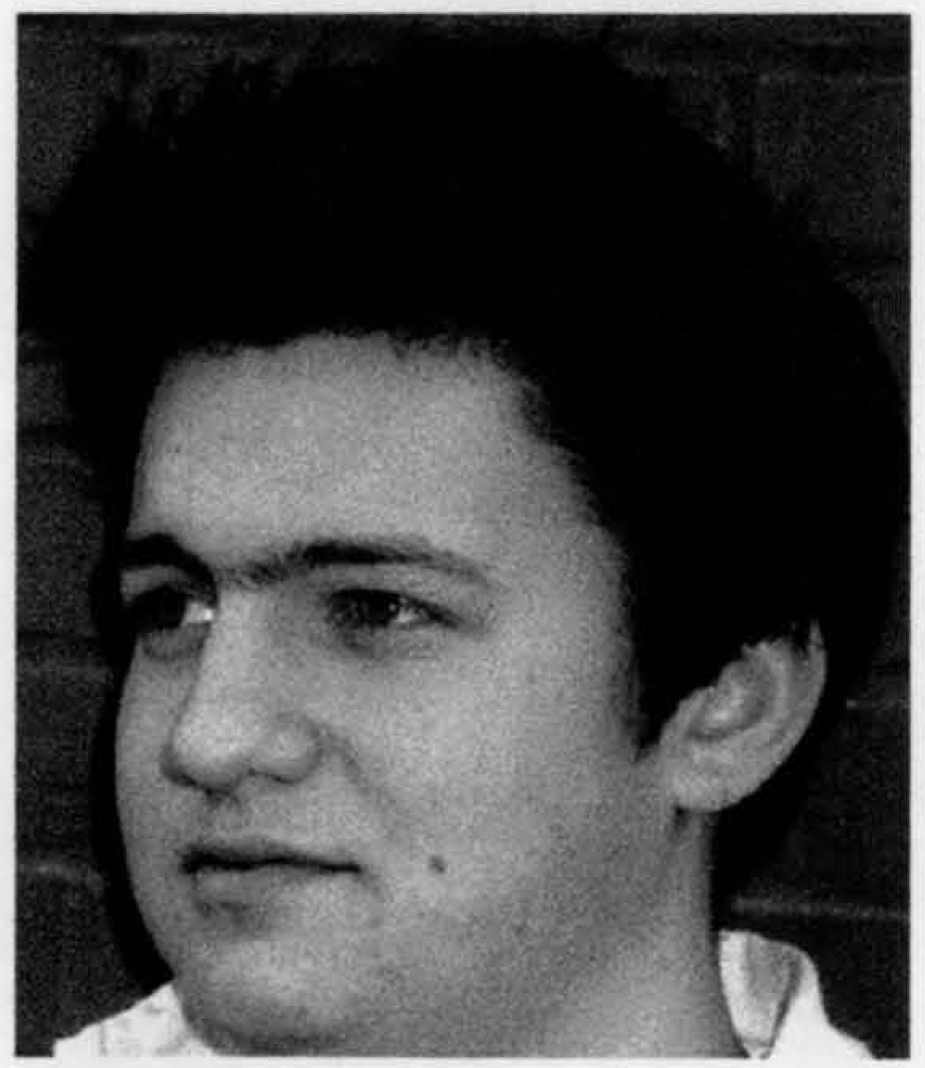
*Actor 19*



*Actor 21*



*Actor 16*

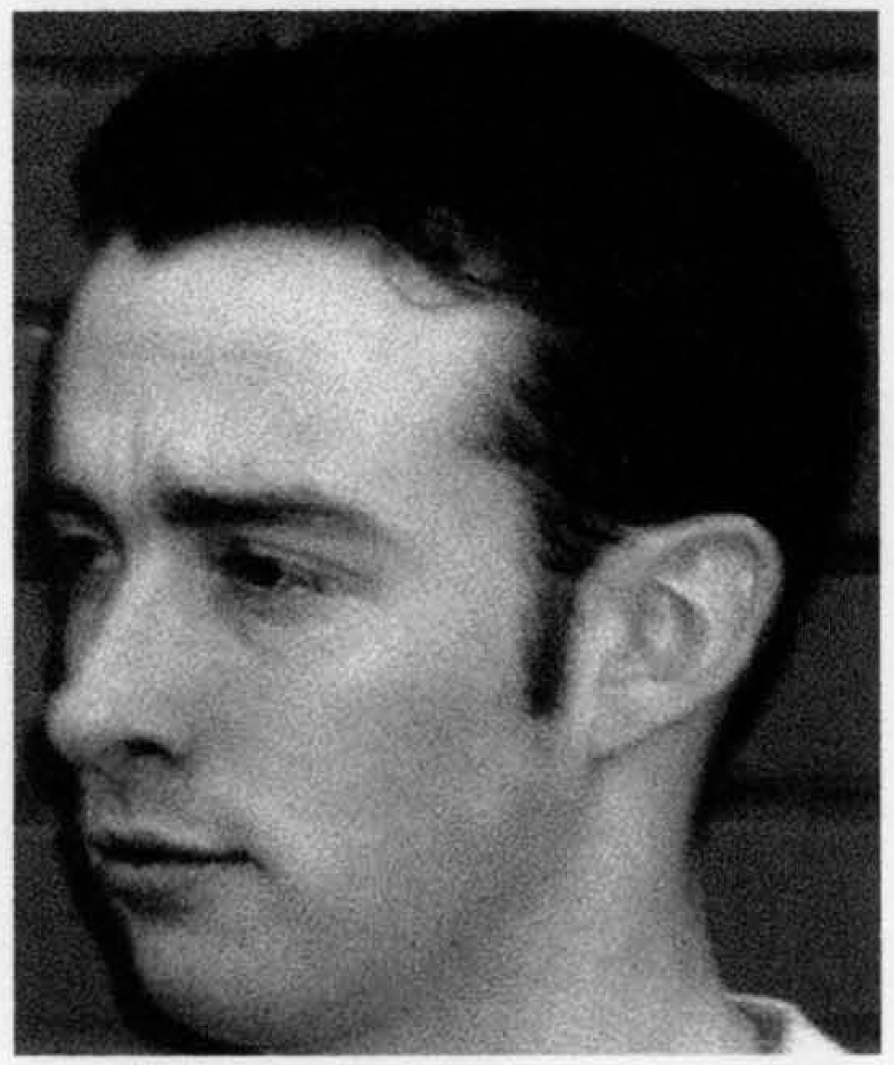


*Actor 18*

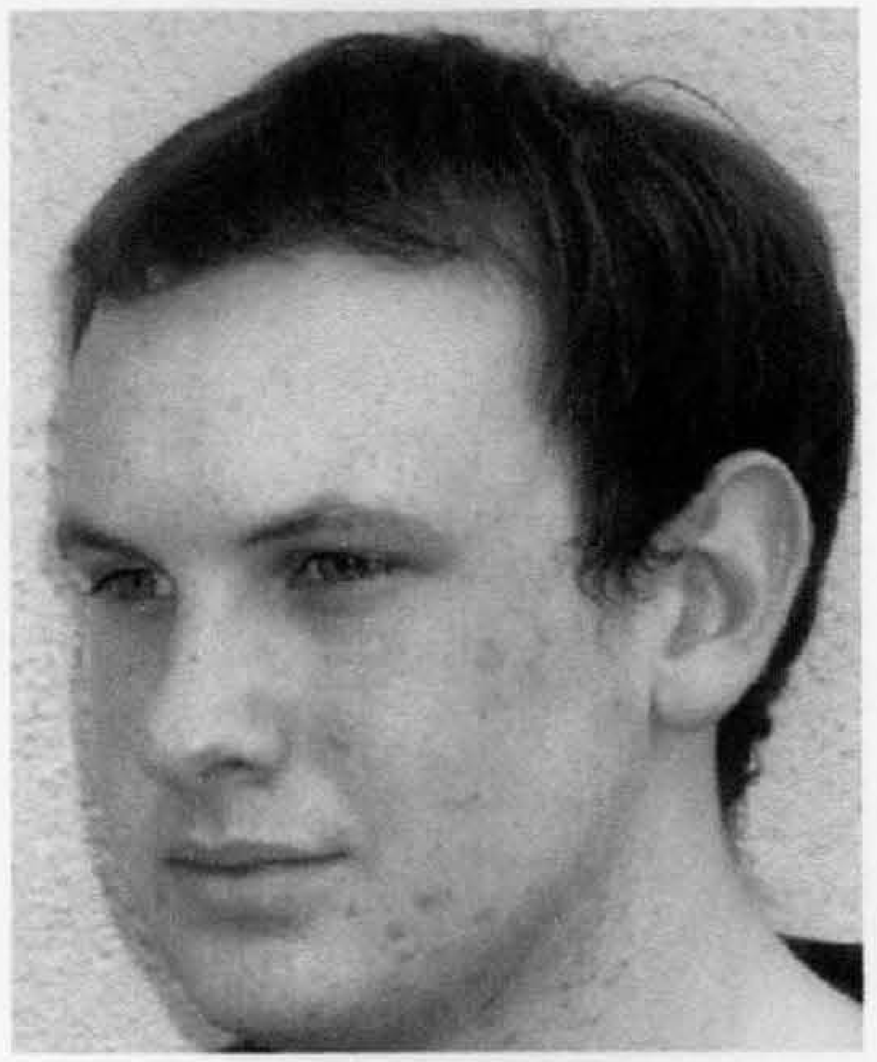
*Actor 20*

*Figure A3: Photographs used in sub-group array number 3 in Experiment 4.1. Actors 15, 16, 17 & 18 were depicted in video in target present conditions. Actor 21 replaced the target for target absent conditions. See section 4.1.1.2 for details.*

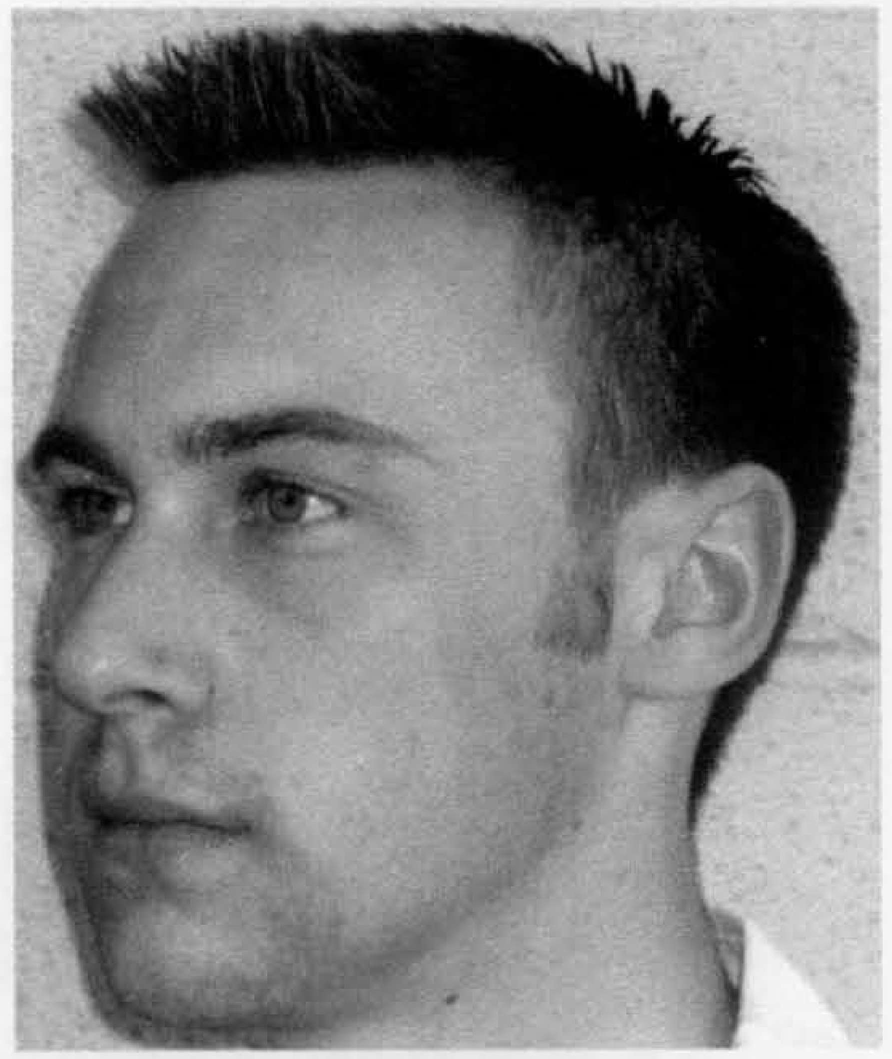
## Appendix A.4: Three-quarters facial portraits used in Experiment 4.1 (sub-group 4)



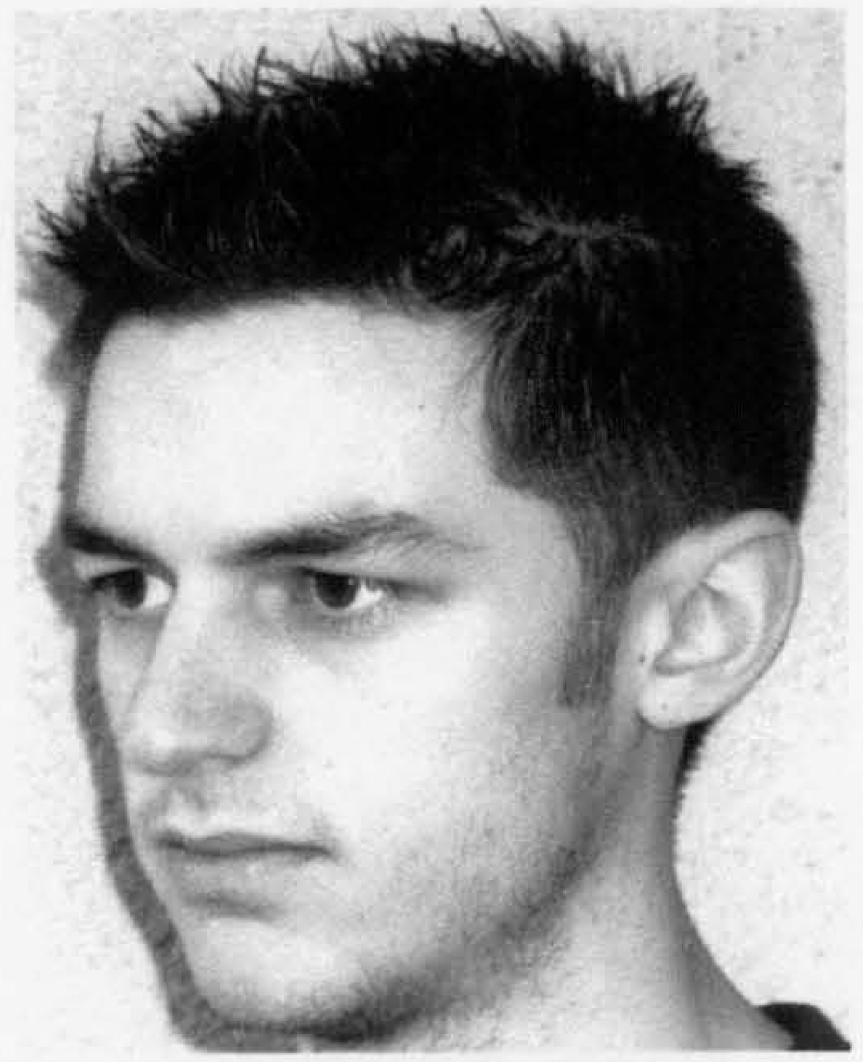
Actor 22



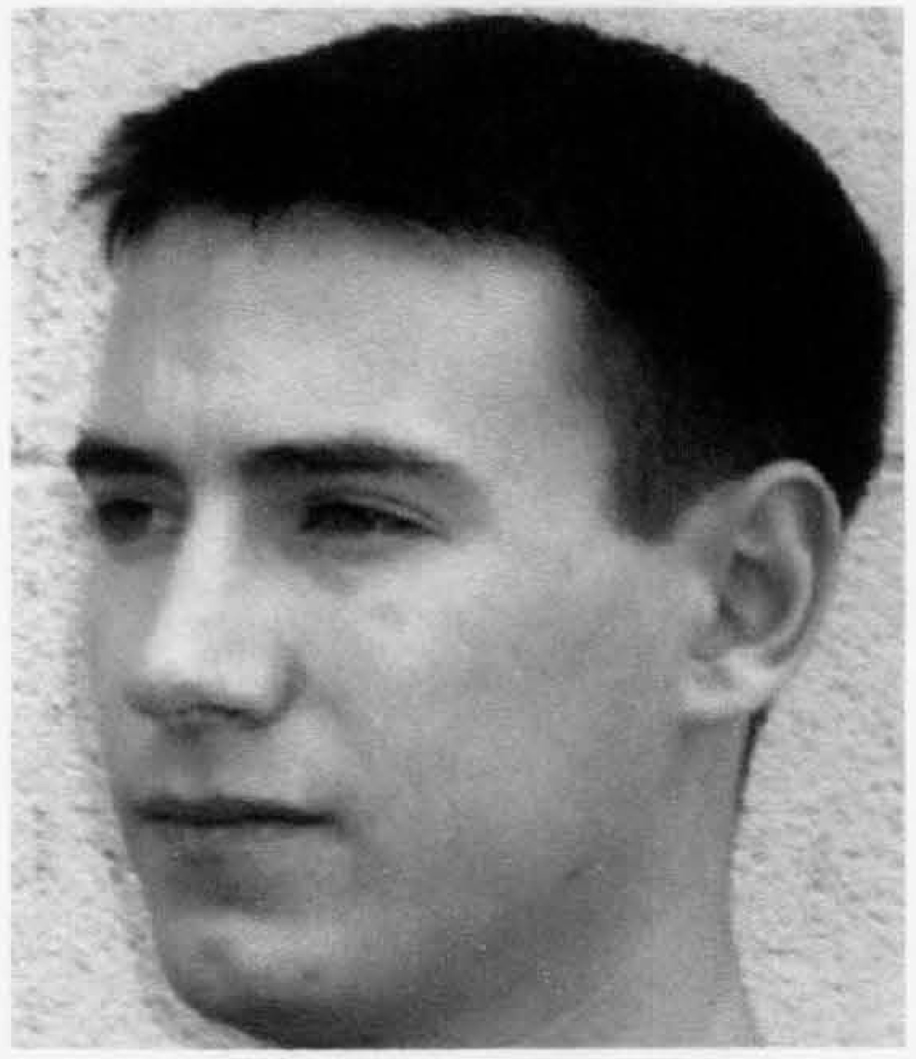
Actor 24



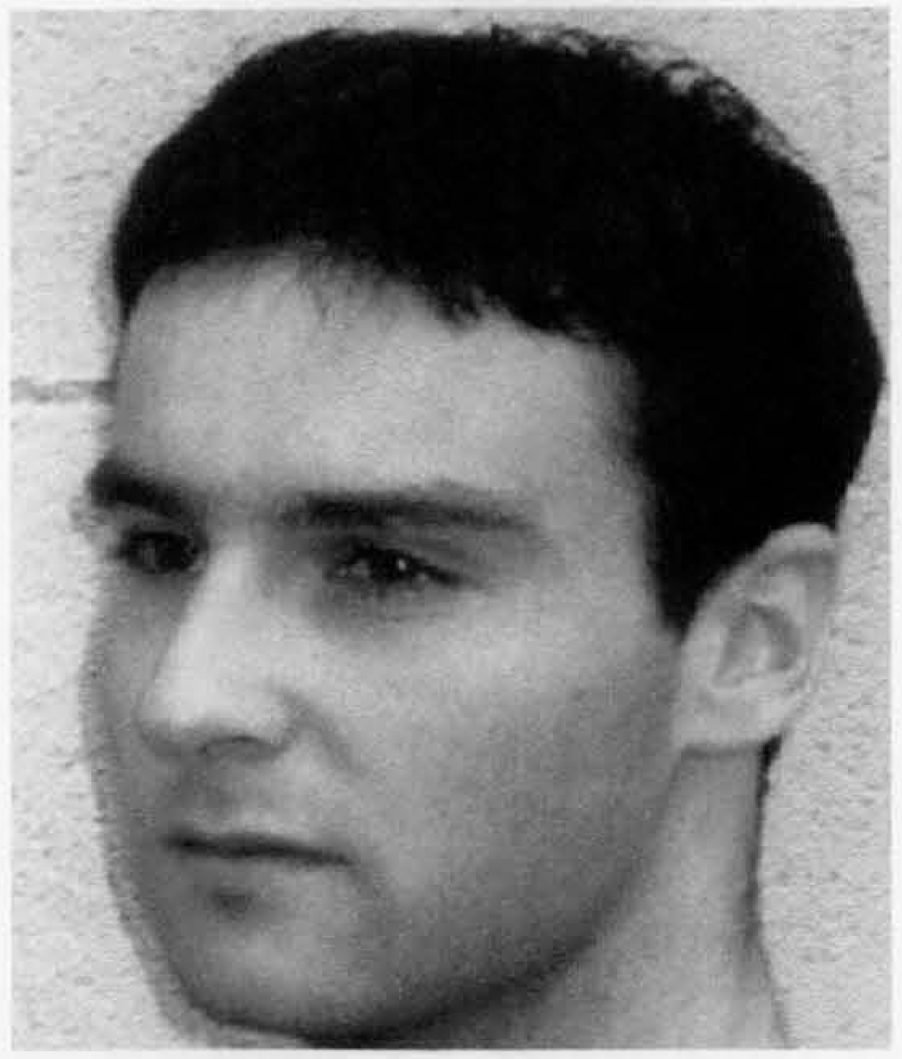
Actor 26



Actor 28



Actor 23

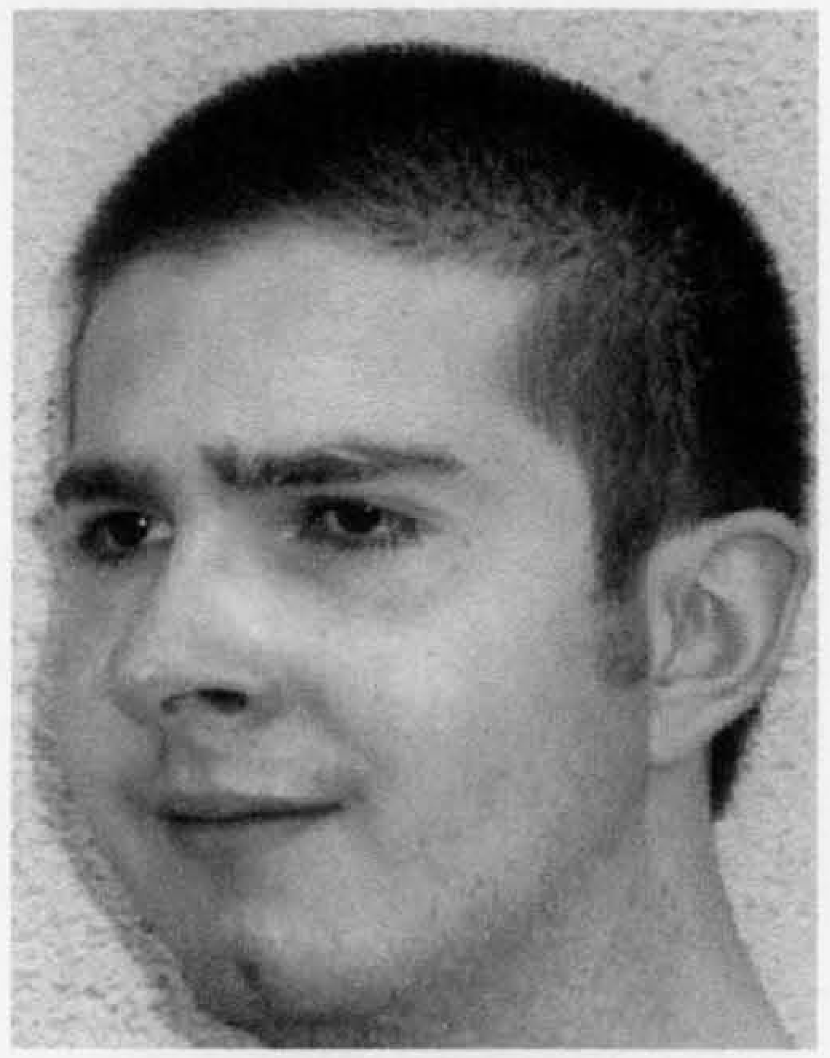


Actor 25

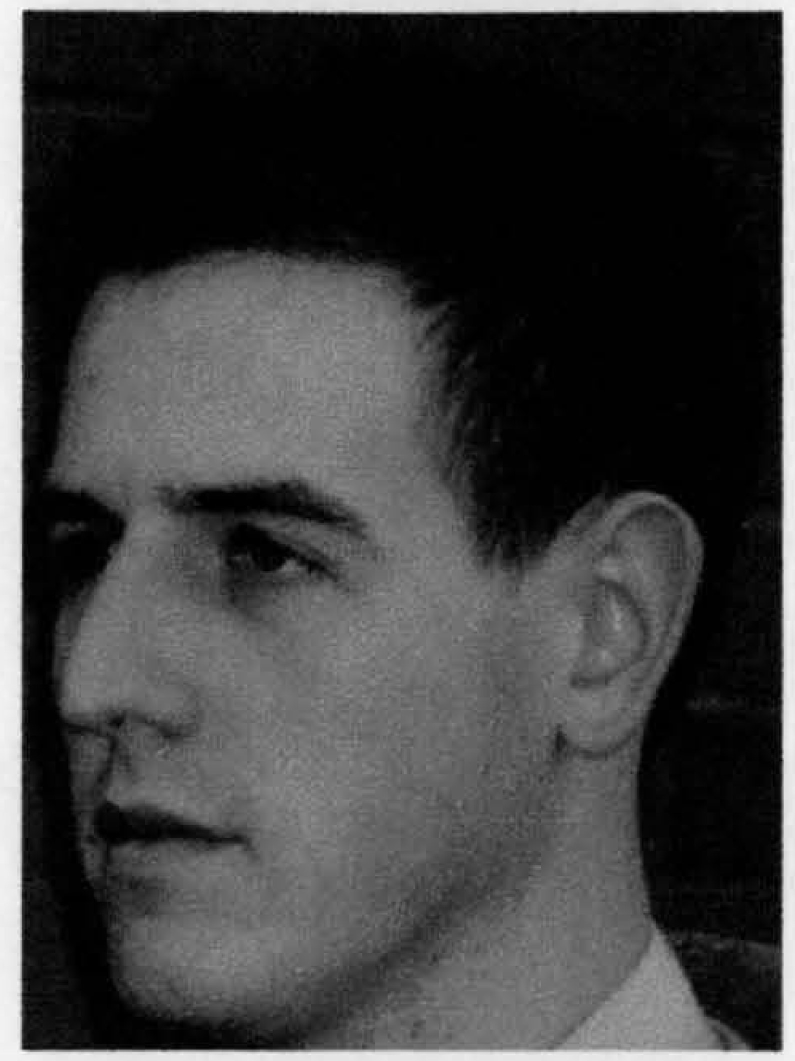
Actor 27

*Figure A4: Photographs used in sub-group array number 4 in Experiment 4.1. Actors 22, 23, 24 & 25 were depicted in video in target present conditions. Actor 28 replaced the target for target absent conditions. See section 4.1.1.2 for details.*

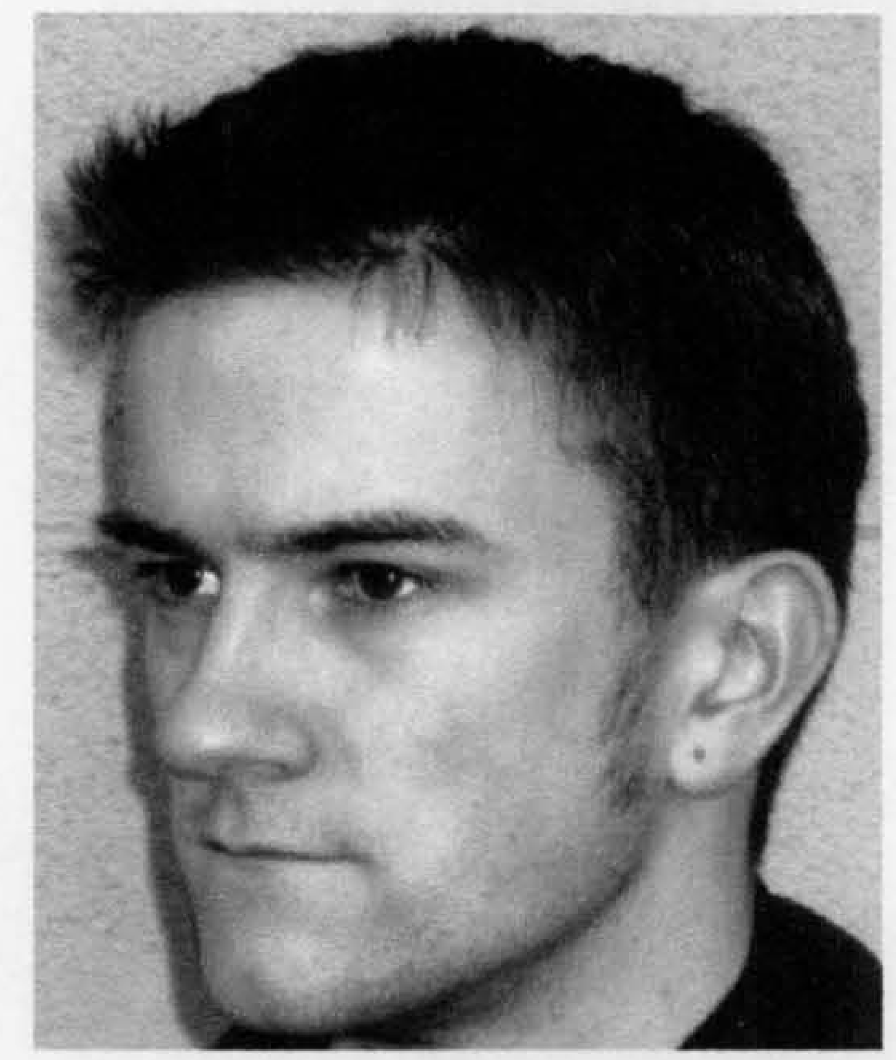
**Appendix A.5: Three-quarters facial portraits used in Experiment 4.1 (sub-group 5)**



*Actor 29*

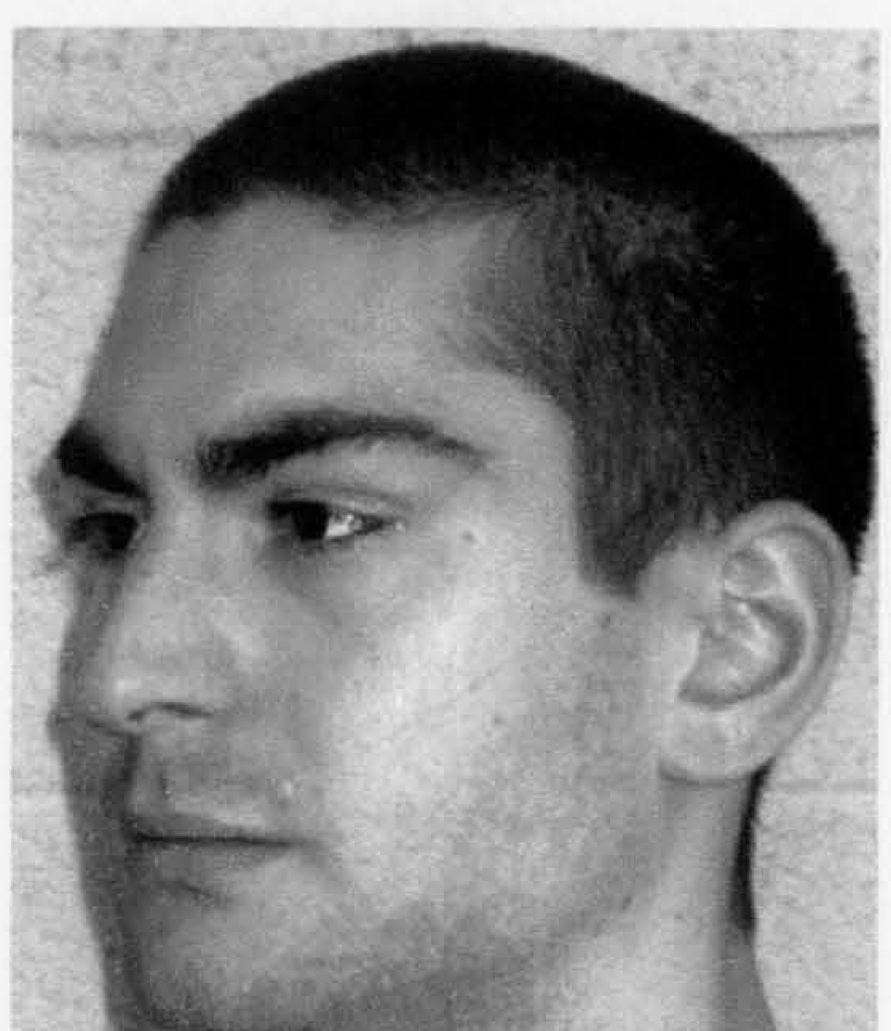


*Actor 31*

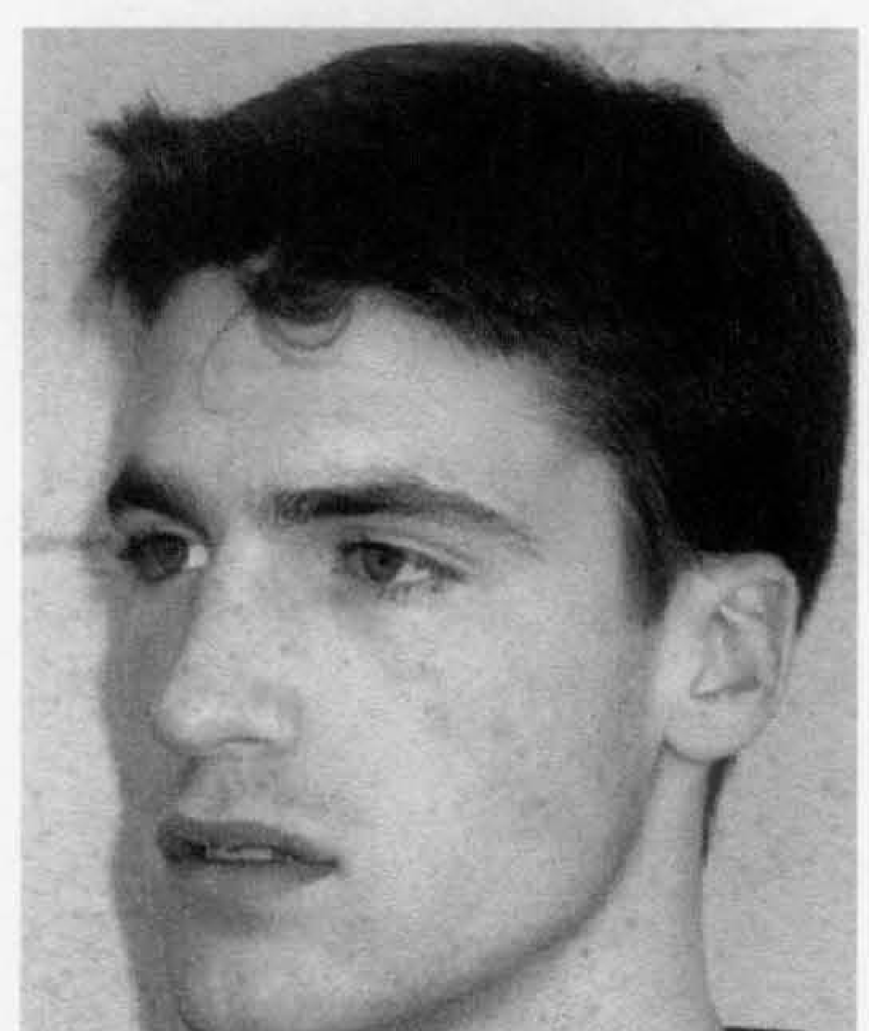


*Actor 33*

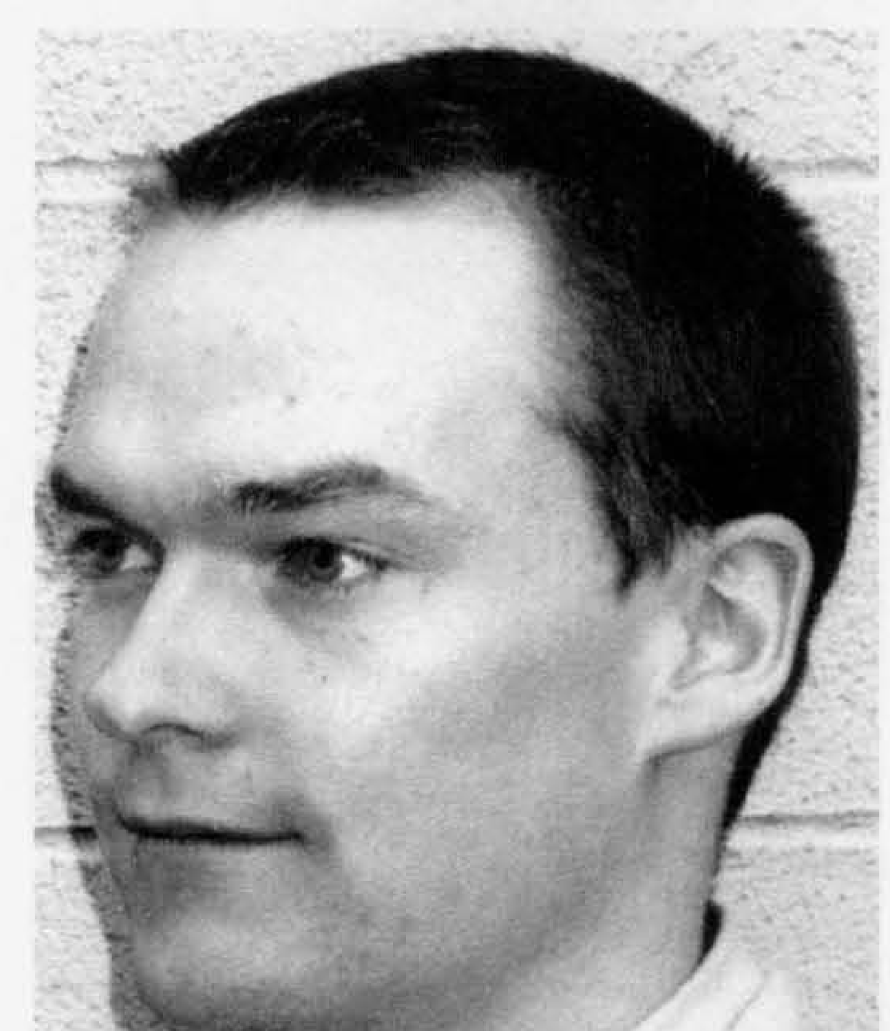
*Actor 35*



*Actor 30*



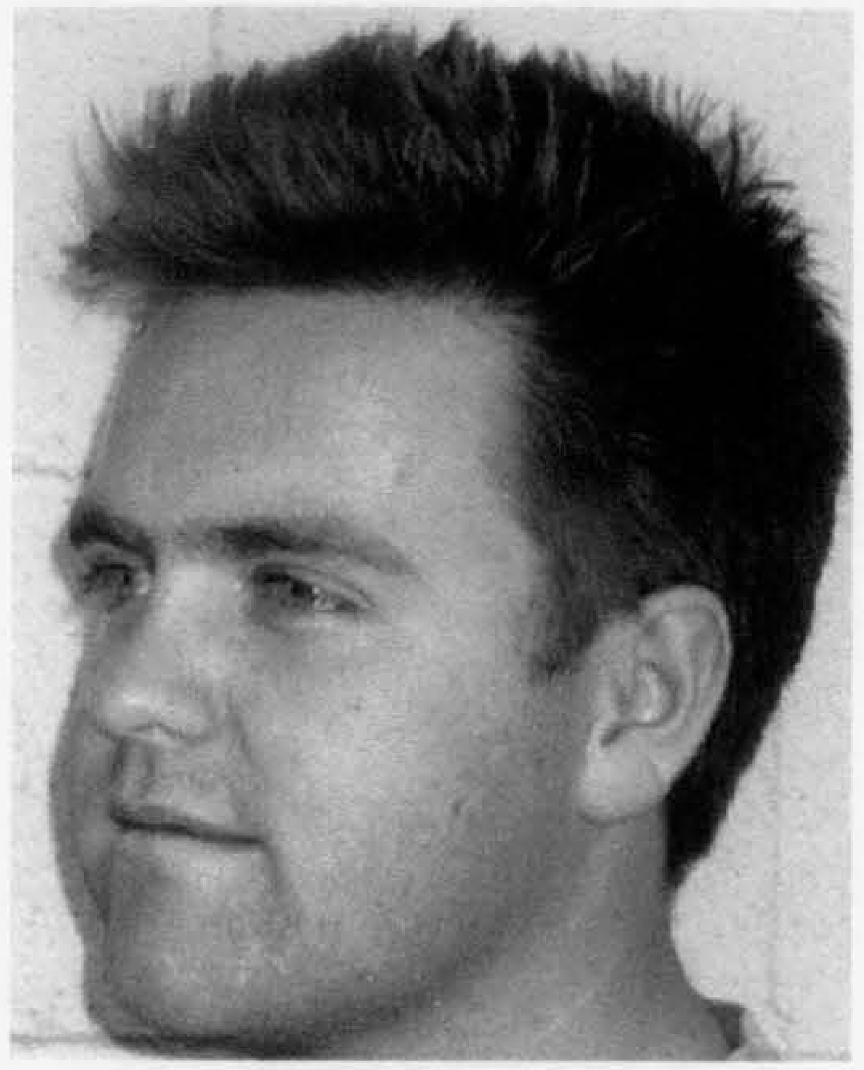
*Actor 32*



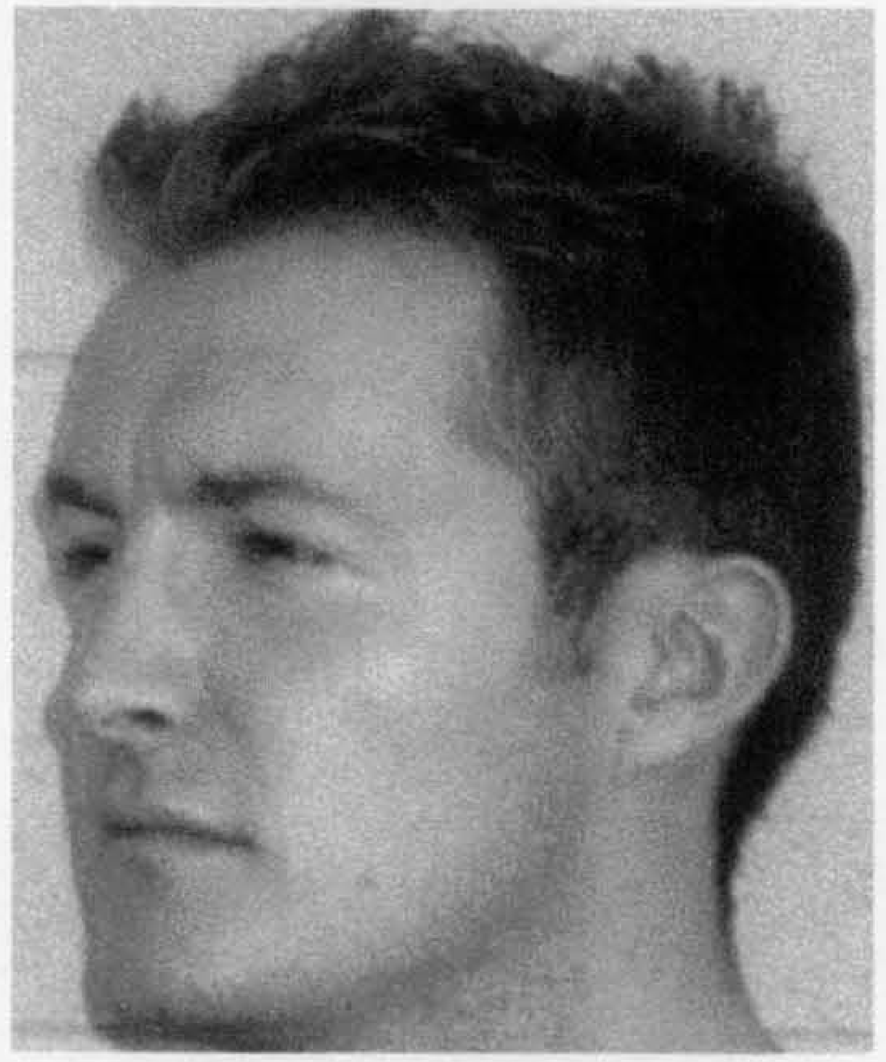
*Actor 34*

*Figure A5: Photographs used in sub-group array number 5 in Experiment 4.1. Actors 29, 30, 31 & 32 were depicted in video in target present conditions. Actor 35 replaced the target for target absent conditions. See section 4.1.1.2 for details.*

## Appendix A.6: Three-quarters facial portraits used in Experiment 4.1 (sub-group 6)

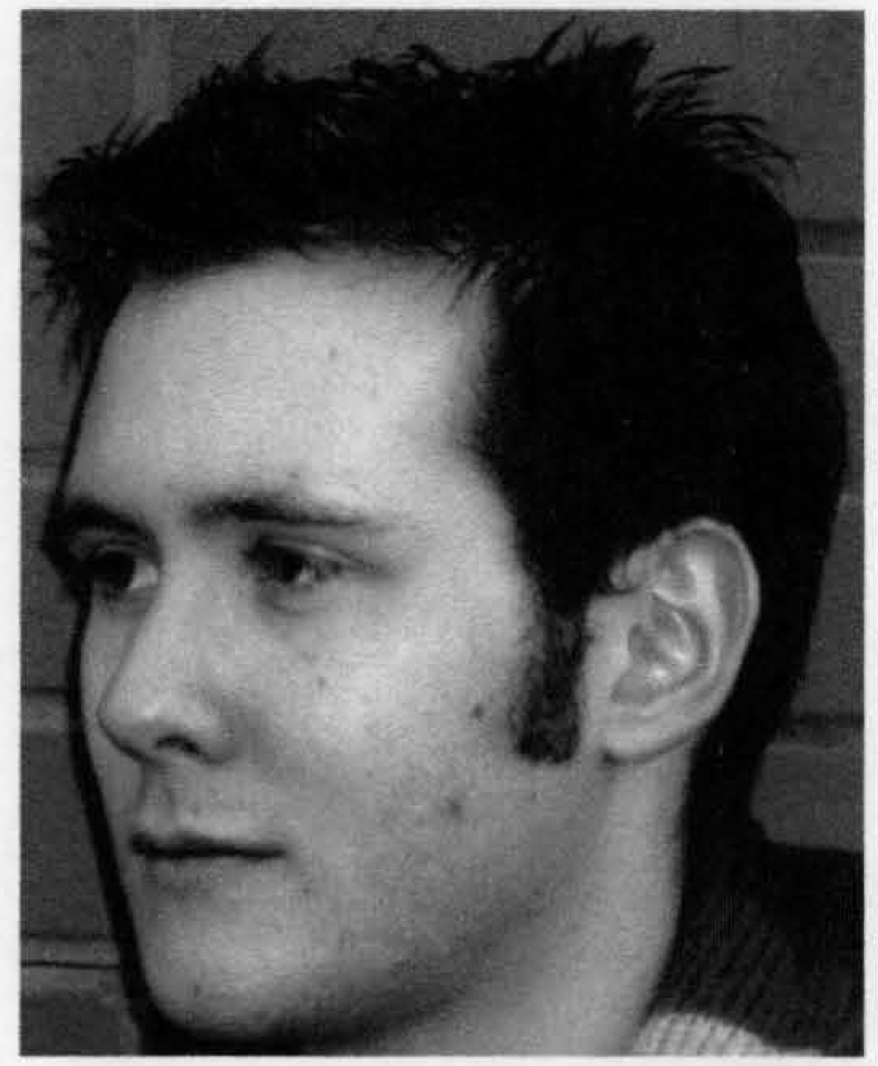


*Actor 36*



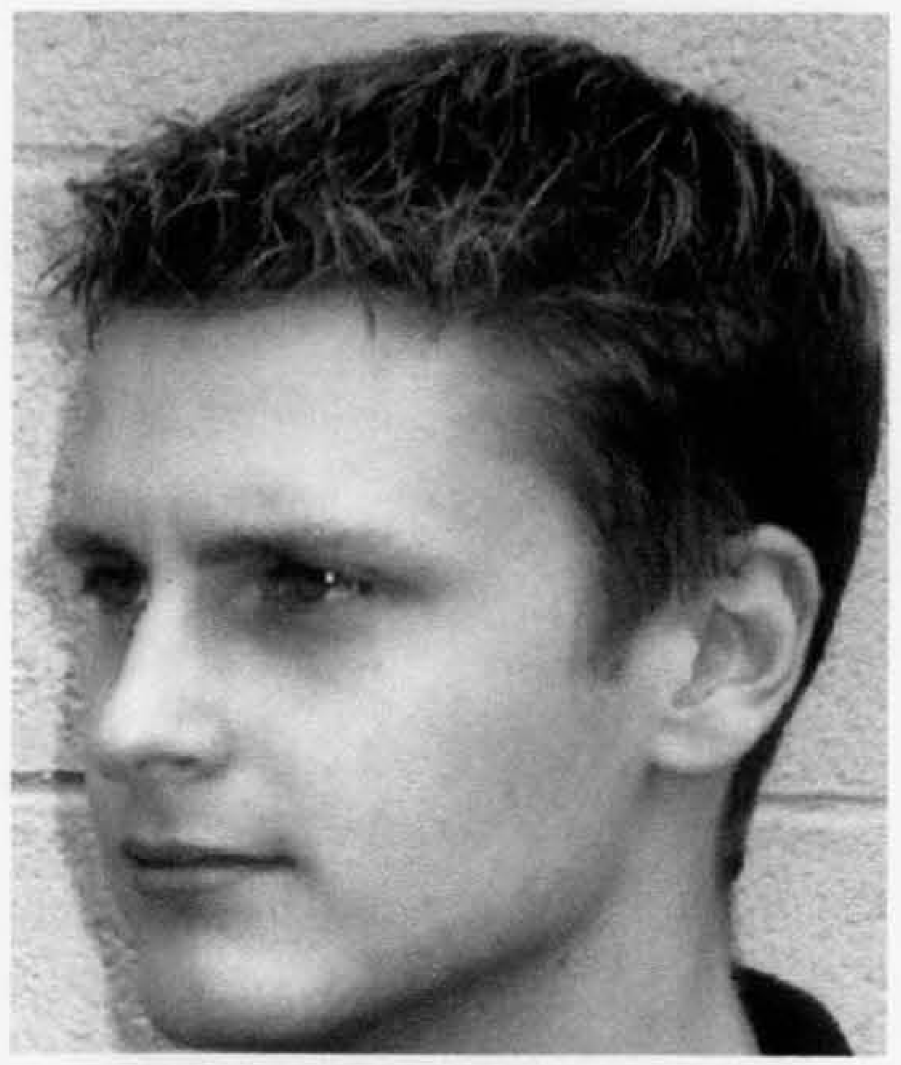
*Actor 38*

*Actor 40*

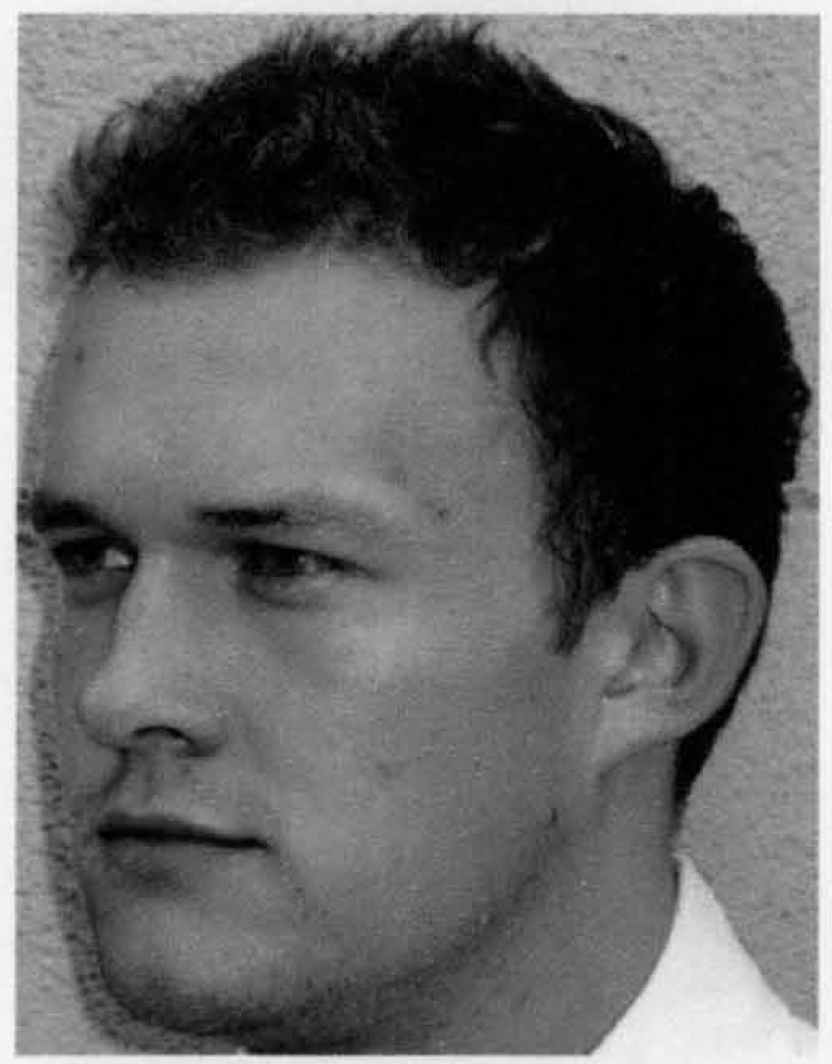


*Actor 42*

*Actor 37*



*Actor 39*



*Actor 41*

*Figure A6: Photographs used in sub-group array number 6 in Experiment 4.1. Actors 36, 37, 38 & 39 were depicted in video in target present conditions. Actor 42 replaced the target for target absent conditions. See section 4.1.1.2 for details.*

## Appendix B.1: Stills from the video footage taken of Actor 43 and Distracter 44 for Experiment 7.1

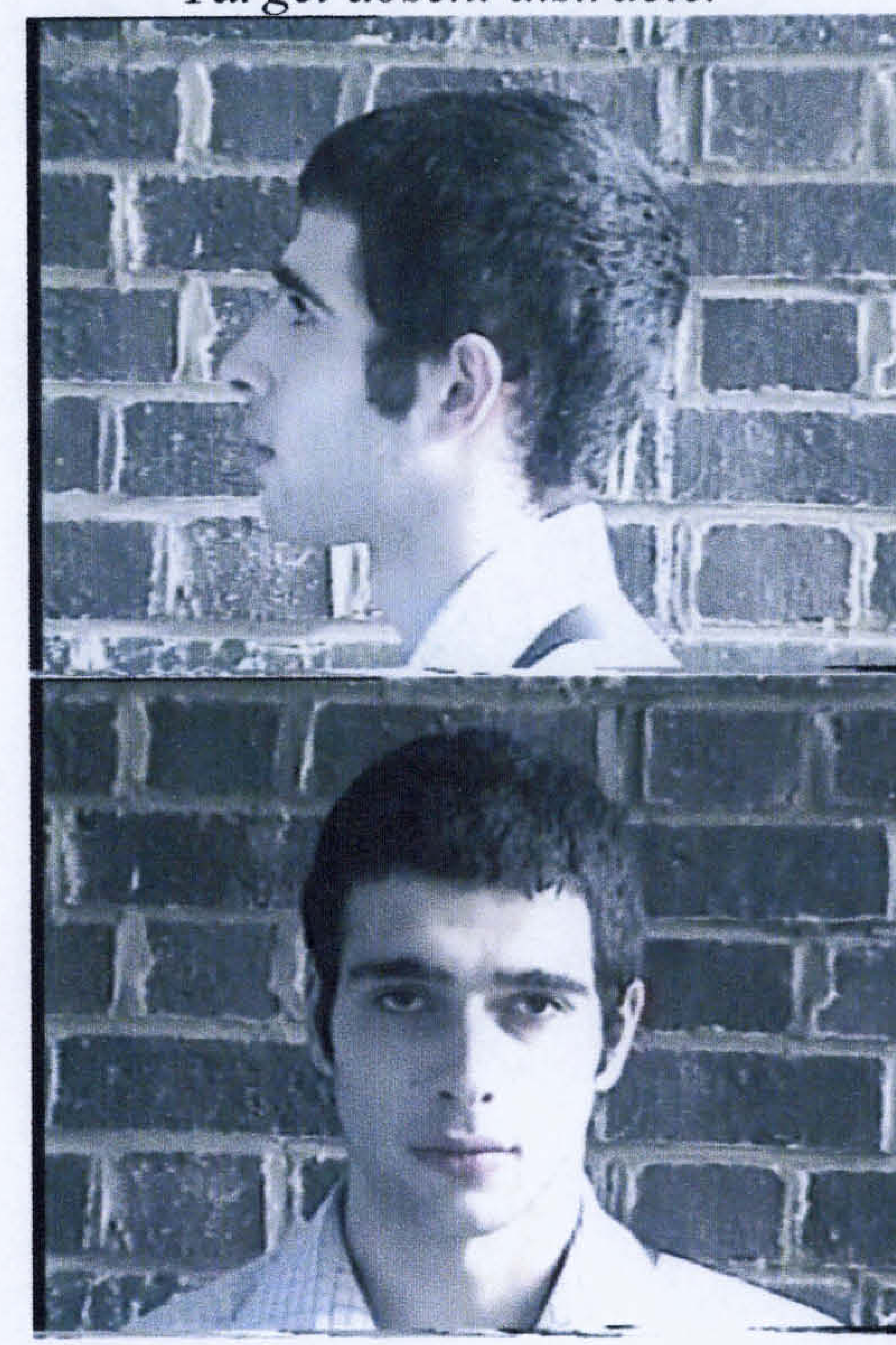
*Target present immediate*

*Actor 43*

*Target present time lapse*

*Actor 43*

*Target absent distracter*

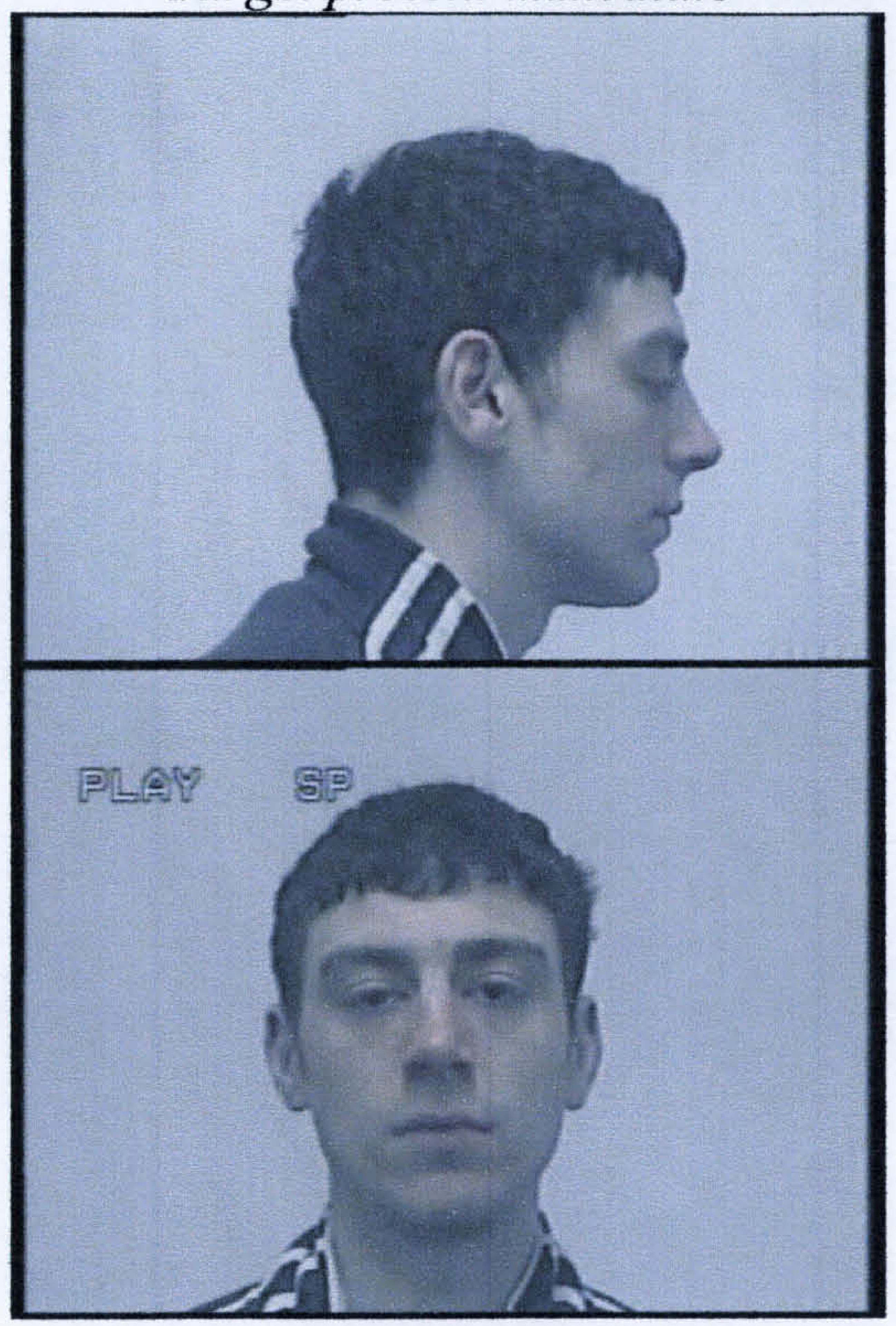


*Distracter 44*

*Figure B.1: Stills from the close-up video footage featuring Actor 43 (left & centre) and Distracter 44 (right): See Section 7.1.2.2 for full details*

**Appendix B.2: Stills from the video footage taken of Actor 46 and Distracter 45 for Experiment 7.1**

*Target present immediate*




*Actor 46*

*Target present time lapse*

*Actor 46*

*Target absent distracter*

*Distracter 45*

*Figure B.2: Stills from the close-up video footage featuring Distracter 45 (right) and Actor 46 (left & centre): See Section 7.1.2.2 for full details*

**Appendix C.1: Details of *DigitalFace* distance and angular measurements**

The following series of figures and tables list the distances automatically calculated by DigitalFace once digitization has been completed. Full details are provided in Chapter 10.

***Table C.1.1: Key to anterior view permanent and transient distances automatically calculated by DigitalFace***

| | | | |
|---|---|---|---|
| Permanent horizontal distances – *see Figure B.1* | | | |
| F1 | Left to right exocanthian distance (1 – 2) | F2 | Inter-pupil distance (3 – 4) |
| F3 | Left to right endocanthian distance (5 – 6) | F4 | Left to right postaurale distance (7 – 8) |
| F5 | Left to right alare distance (9 – 10) | F6 | Left to right cheilion distance (11 – 12) |
| F7 | Face width at height of superior labiale (14 – 15) | F8 | Face width at height of subalare (17 – 18) |
| Permanent vertical distances – *see Figure B.1* | | | |
| F9 | Endocanthian line - subnasale (1 & 2 – 19) | F10 | Endocanthian line – superior labiale (1 & 2 – 13) |
| F11 | Endocanthian line – Stomion (1 & 2 – 20) | F12 | Endocanthian line – inferior labiale (1 & 2 – 21) |
| F13 | Endocanthian line - gnathion (1 & 2 – 22) | F14 | Right superaurale - subaurale (23 - 24) |
| F15 | Left superaurale - subaurale ( 25 - 26) | | |
| Transient distances – *see Figure B.2* | | | |
| F16 | Endocanthian line – trichion (vertical distance 1 & 2 – 27) | F17 | Face width at superciliare height (vertical distance 29 – 30) |
| F18 | Right superciliare – right orbitale superious (vertical distance 28 - 31) | F19 | Left superciliare – left orbitale superious (vertical distance 32 - 33) |
| F20 | Right inner eyebrow – right frontotemporale (horizontal distance 34 - 35) | F21 | Right inner eyebrow – right frontotemporale (horizontal distance 36 - 37) |
| F22 | Distance between inner eyebrows (horizontal distance 34 - 36) | F23 | Endocanthian line - right orbitale superious (vertical distance 1 & 2 – 31) |
| F24 | Endocanthian line - left orbitale superious (vertical distance 1 & 2 – 33) | F25 | Endocanthian line - vertex (vertical distance 1 & 2 – 38) |

*Note: The numbers in the table refer to landmark sites as listed in Figures 10.3*

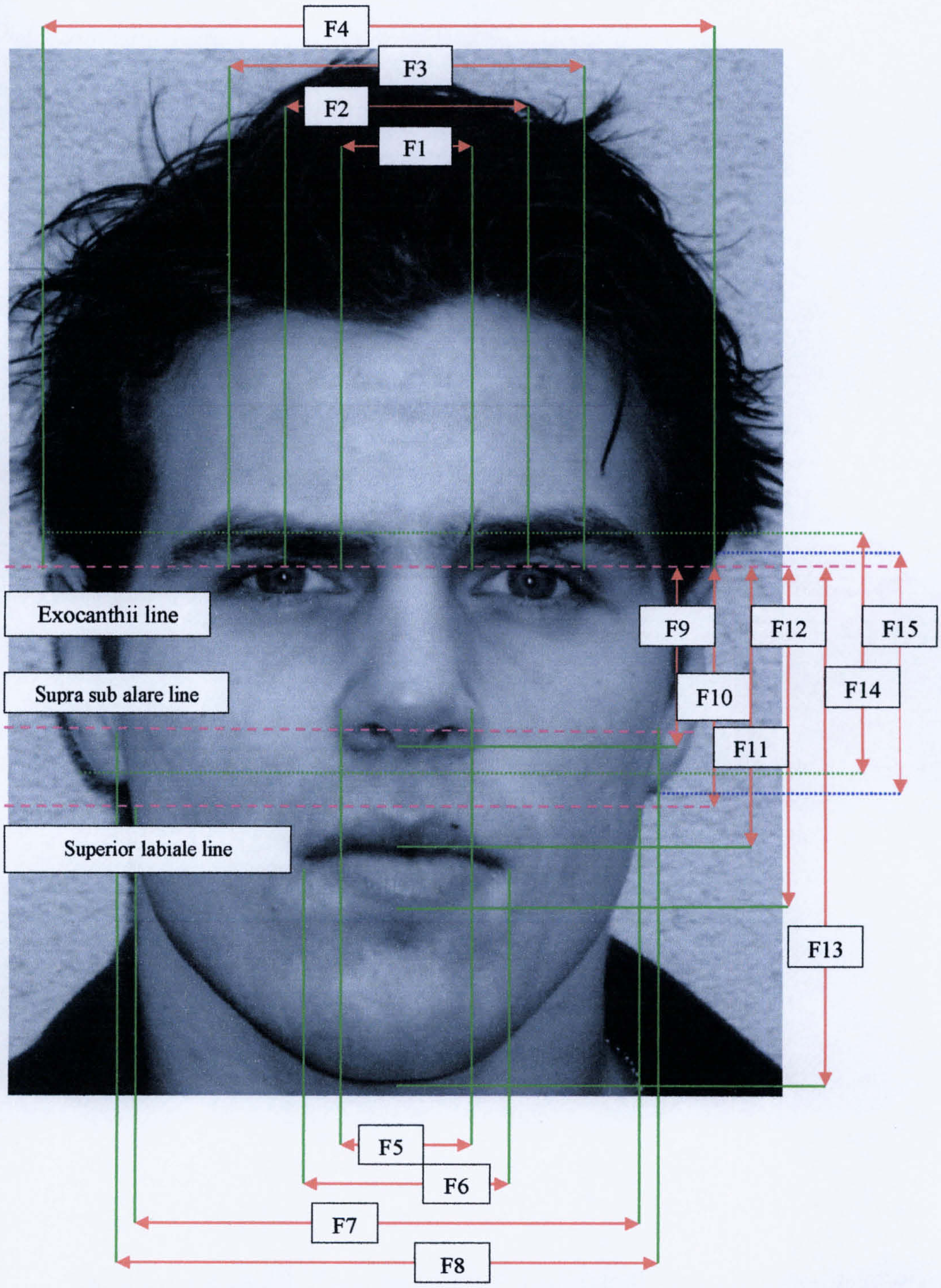



***Figure C.1: Permanent distances measured in anterior view (see Table C.1.1 for key to locations)***

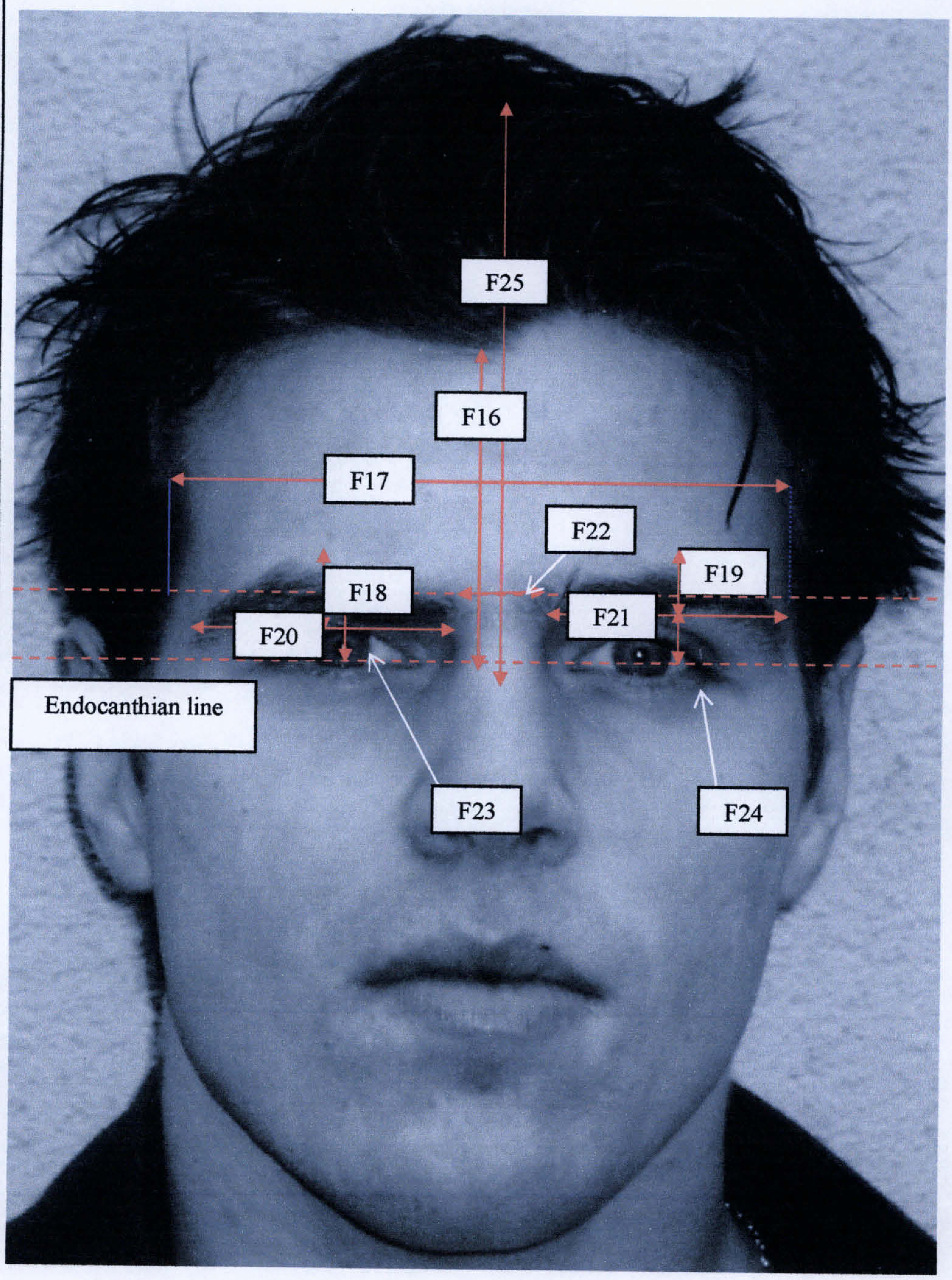



***Figure C.2: Transient distances measured in anterior view (see Table C.1.1 for key to locations)***

***Table C.1.2: Key to anterior angular measurements automatically computed by DigitalFace (see Figure C.3 for locations)***

| | | |
|---|---|---|
| a | *Line 2* | Right endocanthian (1) – Right subaurale (24) |
| | *Line 5* | Bottom of chin (22) - Right subaurale (24) |
| b | *Line 3* | Right exocanthian (5) - Right cheilion (11) |
| | *Line 4* | Right exocanthian (5) – Subnasale (19) |
| c | *Line 1* | Right outside mouth (11) – Right superaurale (23) |
| | *Line 3* | Right exocanthian (5) - Right cheilion (11) |
| d | *Line 2* | Right endocanthian (1) - Right subaurale (24) |
| | *Line 7* | Right endocanthian (1) - Right cheilion (11) |
| e | *Line 5* | Bottom of chin (22) - Right subaurale (24) |
| | *Line 6* | Right outside mouth (11) - Gnathion (22) |
| f | *Line 3* | Right outer canthus (5) – Right cheilion (11) |
| | *Line 7* | Right inner canthus (1) - Right cheilion (11) |
| g | *Line 6* | Right cheilion (11) - Gnathion (22) |
| | *Line 8* | Left cheilion (12) - Gnathion (22) |
| h | *Line 4* | Right exocanthian - Subnasale (5 – 19) |
| | *Line 10* | Left exocanthian - Subnasale (6 – 19) |
| i | *Line 9* | Left endocanthian (2) - Left cheilion (12) |
| | *Line 13* | Left endocanthian (2) - Left subaurale (26) |
| j | *Line 9* | Left endocanthian (2) - Left cheilion (12) |
| | *Line 11* | Left exocanthian (6) - Left cheilion (12) |
| k | *Line 10* | Left exocanthian (6) - Subnasale (19) |
| | *Line 11* | Left exocanthian (6) - Left cheilion (12) |
| l | *Line 11* | Left exocanthian (6) - Left cheilion (12) |
| | *Line 14* | Left cheilion (12) – Left superaurale (25) |
| m | *Line 12* | Gnathion (22) - Left subaurale (26) |
| | *Line 13* | Left endocanthian (2) - Left subaurale (26) |
| n | *Line 8* | Left cheilion (12) - Gnathion (22) |
| | *Line 12* | Gnathion (22) - Left subaurale (26) |

*Note: Numbers indicate landmark sites as listed in Figure 10.3*

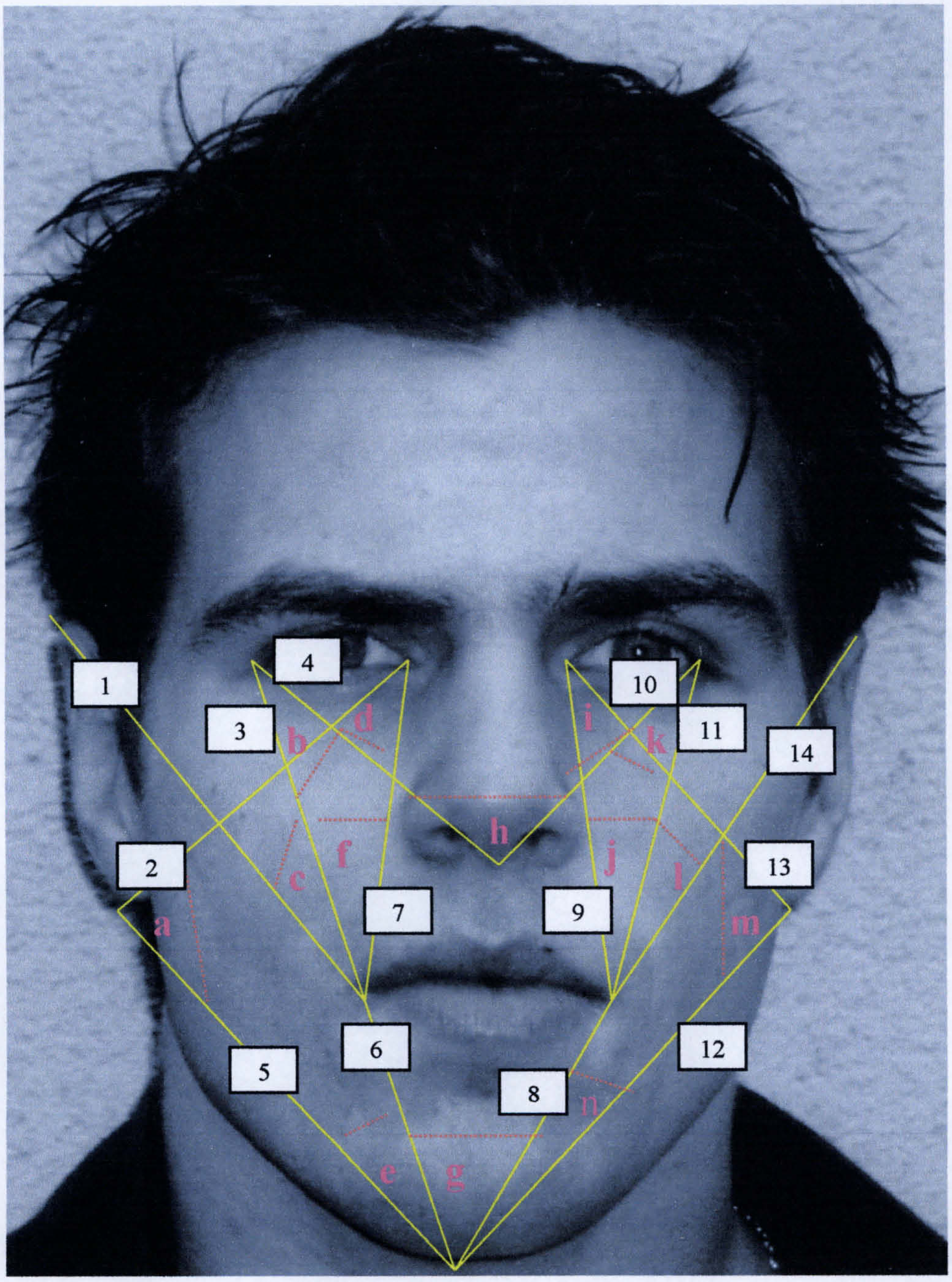



***Figure C.3: Angular measurements in anterior view automatically calculated by DigitalFace (see Table C.1.2 for key to locations)***

***Table C.1.3: Key to permanent and transient profile distances automatically calculated by DigitalFace (see Figure C.4 for locations)***

| | | | |
|---|---|---|---|
| Permanent vertical distances | | | |
| P1 | Subnasale/subaurale baseline – Pronasale (1 & 2 – 3) | P2 | Subnasale/subaurale baseline – Endocanthian (1 & 2 – 4) |
| P3 | Subnasale/subaurale baseline – Superaurale (1 & 2 – 5) | P4 | Subnasale/subaurale baseline – Cheilion (1 & 2 – 6) |
| P5 | Subnasale/subaurale baseline – Gnathion (1 & 2 – 7) | | |
| Permanent horizontal distances | | | |
| P6 | Postaurale – otobasion infrious (8 – 9) | P7 | Postaurale – Endocanthian (8 – 4) |
| P8 | Postaurale – Right alare curvature (8 – 10) | P9 | Postaurale – Pronasale (8 – 3) |
| P10 | Pronasale – Sellion (3 – 11) | | |
| Transient distances | | | |
| P11 | Glabella – Frontotemporale (12 – 13) | P12 | Subnasale/subaurale baseline – vertex (1 & 2 – 14) |

*Note: Numbers indicate landmark sites as listed in Figure 10.4*

***Table C.1.4: Key to profile angular measurements automatically computed by DigitalFace (see Figure C.5 for locations)***

| Angle | Lines | |
|---|---|---|
| A | Line X | Pronasale (3) – Sellion (11) |
| | Line Z | Pronasale (3) - Gnathion (7) |
| B | Line V | Subaurale (2) - pronasale (3) |
| | Line Z | Pronasale (3) - Gnathion (7) |
| C | Line X | Pronasale (3) – Sellion (11) |
| | Line V | Subaurale (2) - Pronasale (3) |
| D | Line U | Alar Curvature (10) – Sellion (11) |
| | Line X | Pronasale (3) – Sellion (11) |
| E | Line U | Alar Curvature (10) – Sellion (11) |
| | Line Y | Subaurale (2) – Exocanthian (4) |
| F | Line X | Pronasale (3) – Sellion (11) |
| | Line Y | Subaurale (2) – Exocanthian (4) |
| G | Line T | Subaurale (2) – Superaurale (5) |
| | Line X | Pronasale (3) – Sellion (11) |
| H | Line T | Subaurale (2) – Superaurale (5) |
| | Line Y | Subaurale (2) – Exocanthian (4) |
| I | Line V | Subaurale (2) - Pronasale (3) |
| | Line Y | Subaurale (2) – Exocanthian (4) |
| J | Line S | Subaurale (2) – Gnathion (7) |
| | Line V | Subaurale (2) - Pronasale (3) |
| K | Line S | Subaurale (2) – Gnathion (7) |
| | Line Z | Pronasale (3) - Gnathion (7) |

*Note: Numbers indicate landmark sites as listed in Figure 10.4*

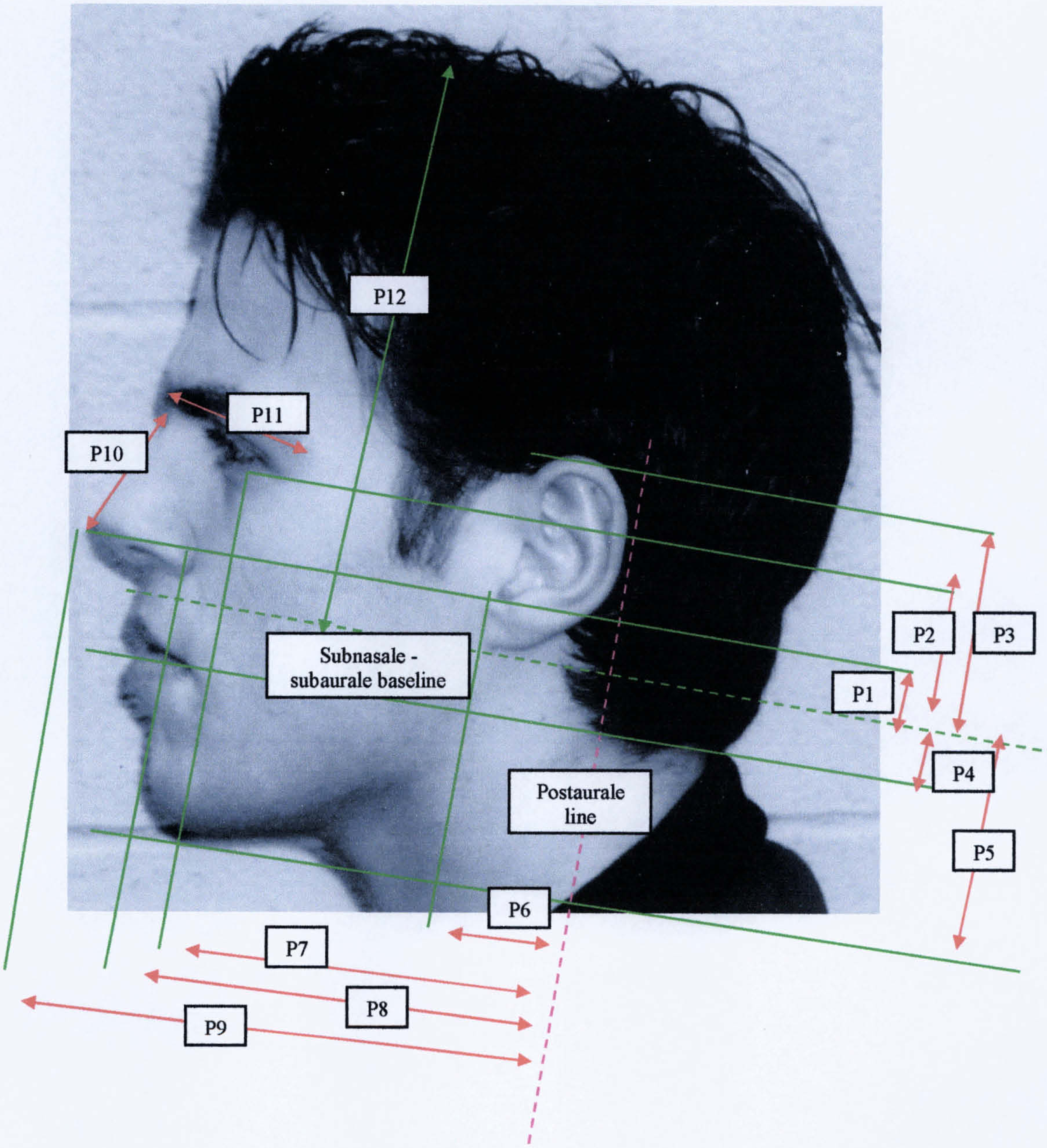



***Figure C.4: Key to distances automatically calculated by DigitalFace in profile view (see Table C.1.3 for key to locations)***

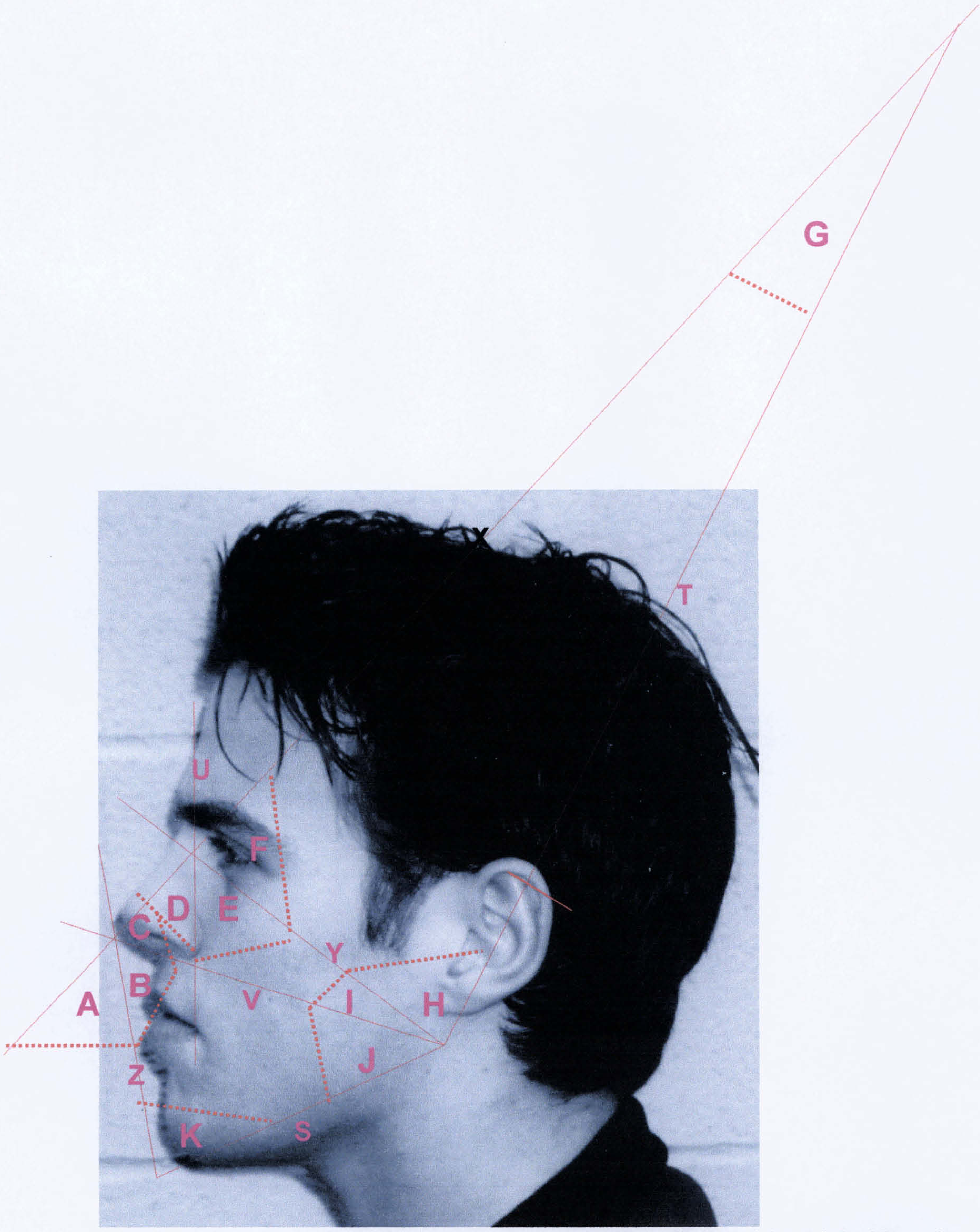



*Figure C.5. Angular measurements automatically calculated by DigitalFace in profile view (see Table C.1.4 for key to locations)*

## Appendix C.2

The following series of tables and figures lists the mean and standard deviations for all final converted unique independent distance and angular proportional vectors used within all analyses. The values for the anterior view tables were derived from the 100 database faces. The profile view tables were derived from the 70 database faces. Full details are provided in Chapter 10. Numbers indicate landmark sites listed in Figures 10.3 and 10.4.

***Table C.2.1: Mean permanent anterior horizontal ratios for converted distance values***

| Ratio Measure | Permanent anterior horizontal distances | | Mean ratios (SD in parentheses) |
|---|---|---|---|
| FH1 | F1 | Distance between exocanthian (1) – (2) | .191 (.016) |
| FH2 | F5 – F1 | Distance between alare (9 – 10) minus distance between exocanthian (1 – 2) | .029 (.018) |
| FH3 | F6 – F5 | Distance between cheilion (11 – 12) minus distance between alare ( 9 – 10) | .078 (.024) |
| FH4 | F2 – F6 | Distance between pupils (3 – 4) minus distance between cheilion (11 – 12) | .071 (.027) |
| FH5 | F3 – F2 | Distance between endocanthian (5 – 6) minus distance between pupils (3 – 4) | .170 (.014) |
| FH6 | F7 – F3 | Face width at height of superior labiale (14 – 15) minus distance between endocanthian (5 – 6) | .191 (.045) |
| FH7 | F8 – F7 | Face width at height of subaurale (17 – 18) minus face width at height of superior labiale (14 – 15) | .046 (.017) |
| FH8 | F4 – F8 | Distance between postaurale (7 – 8) minus face width at height of subaurale (17 – 18) | .224 (.046) |

***Table C.2.2: Mean permanent anterior vertical ratios for converted distance values***

| Ratio measure | Permanent anterior vertical distances | | Mean ratios (SD in parentheses) |
|---|---|---|---|
| FV1 | F9 | Distance between endocanthian line (1/2) and subnasale (19) | .389 (.032) |
| FV2 | F10 – F9 | Distance between superior labiale (13) and subnasale (19) | .131 (.019) |
| FV3 | F11– F10 | Distance between stomion (20) and superior labiale (13) | .061 (.010) |
| FV4 | F12– F11 | Distance between inferior labiale (21) and stomion (20) | .092 (.018) |
| FV5 | F13– F12 | Distance between gnathion (22) and inferior labiale (21) | .327 (.033) |
| FV6 | F14 | Distance between right superaurale (23) and subaurale (24) | .478 (.055) |
| FV7 | F15 | Distance between left superaurale (25) and subaurale (26) | .476 (.055) |

***Table C.2.3: Mean transient anterior ratios for converted distance values***

| Ratio Measure | | Transient anterior distances | Mean ratios (SD in parentheses) |
|---|---|---|---|
| FTV1 | F16 | Distance between endocanthian line (1 & 2) and trichion (27) | .659 (.084) |
| FTV2 | F25-F16 | Distance between vertex (38) and trichion (1/2 – 27) | .424 (.120) |
| FT1 | F17 | Face width at height of superciliare (29 - 30) | .772 (.051) |
| FT2 | F18 | Distance between right superciliare (28) and right orbitale superious (31) | .112 (.036) |
| FT3 | F19 | Distance between left superciliare (32) and left orbitale superious (33) | .107 (.030) |
| FT4 | F20 | Distance between right inner eyebrow (34) and right frontotemporale (35) | .245 (.033) |
| FT5 | F21 | Distance between right inner eyebrow (36) and right frontotemporale (37) | .250 (.035) |
| FT6 | F22 | Distance between left (36) and right inner eyebrows (34) | .179 (.032) |
| FT7 | F23 | Distance between endocanthian line (1 & 2) and right orbitale superious (31) | .080 (.029) |
| FT8 | F24 | Distance between endocanthian line (1 & 2) and left orbitale superious (33) | .081 (.027) |

***Table C.2.4. Mean anterior angular measurements (standard deviations in parentheses; full details of angular measurements provided in Table C.1.2 and Figure C.3)***

| Angle | Mean | Angle | Mean | Angle | Mean |
|---|---|---|---|---|---|
| a | 87.15 (5.85) | b | 28.79 (2.64) | c | 18.15 (4.24) |
| d | 40.88 (5.79) | e | 15.67 (4.79) | f | 24.93 (1.65) |
| g | 54.74 (5.35) | h | 91.16 (5.85) | i | 42.14 (5.46) |
| j | 24.61 (1.77) | k | 28.56 (2.86) | l | 19.11 (4.10) |
| m | 86.26 (5.63) | n | 17.48 (4.94) | | |

*Table C.2.5: Mean permanent profile horizontal ratios for converted distance values*

| Ratio Measure | Permanent profile horizontal distances | | Mean ratios (SD in parentheses) |
|---|---|---|---|
| PH1 | P9 – P8 | Distance between pronasale (3) and alar curvature (10) | .257 (0.017) |
| PH2 | P8 – P7 | Distance between exocanthian (4) and right alar curvature (10) | .113 (0.029) |
| PH3 | P7 – P6 | Distance between otobasion infrious (9) and exocanthian (4) | .630 (.0281) |
| PH4 | P6 | Distance between postaurale (8) and otobasion infrious (9) | .279 (.0312) |

*Table C.2.6: Mean permanent profile vertical ratios for converted distance values*

| Ratio measure | Permanent profile vertical distances | | Mean ratios (SD in parentheses) |
|---|---|---|---|
| PV1 | P5 – P4 | Distance between gnathion (7) and cheilion (6) | .192 (.0224) |
| PV2 | P4 | Distance between cheilion (6) and subnasale to subaurale line (1 – 2) | .091 (.0169) |
| PV3 | P1 | Distance between subnasale to subaurale line (1 – 2) and pronasale (3) | .312 (.0267) |
| PV4 | P2 – P1 | Distance between exocanthian (4) and pronasale (3) | .156 (.0389) |
| PV5 | P3 – P2 | Distance between superaurale (5) and exocanthian (4) | .398 (.027) |
| PV6 | P10 | Distance between pronasale (3) and sellion (11) | 1.009 (.111) |

*Table C.2.7: Mean transient profile ratios for converted distance values*

| Ratio measure | Transient profile distances | | Mean ratios (SD in parentheses) |
|---|---|---|---|
| PVT1 | P12 – P3 | Distance between vertex (14) and superaurale (5) | .228 (.0615) |
| PHT1 | P11 | Distance between glabella (12) and rear frontotemporale (13) | .211 (.0558) |

***Table C.2.8: Mean profile angular measurements (standard deviations in parentheses; full details of angular measurements provided in Table C.1.4 and Figure C.5)***

| Angle | Mean | Angle | Mean | Angle | Mean |
|---|---|---|---|---|---|
| A | 56.42<br>(4.37) | B | 53.98<br>(2.85) | C | 69.63<br>(4.02) |
| D | 35.02<br>(3.18) | E | 50.93<br>(3.83) | F | 94.03<br>(4.56) |
| G | 14.58<br>(5.50) | H | 71.44<br>(5.12) | I | 24.38<br>(2.05) |
| J | 44.90<br>(3.79) | K | 81.11<br>(4.19) | | |

***Table C.2.9: Restricted distance and angular vectors in anterior and profile view for when DigitalFace was applied to disguised images***

***Anterior distance vectors (Tables B.2.1; B.2.2; & B.2.3)***

| | | | | |
|---|---|---|---|---|
| FH1 | FH2 | FH3 | FH4 | FH5 |
| FH6 | FH7 | FT2 | FT3 | FT4 |
| FT5 | FT6 | FT7 | FT8 | FV1 |
| FV2 | FV3 | FV4 | FV5 | |

***Anterior angular vectors (Figure B.3)***

| | | | | | |
|---|---|---|---|---|---|
| b | f | g | h | j | k |

***Profile distance vectors (Tables B.2.5; B.2.6; & B.2.7)***

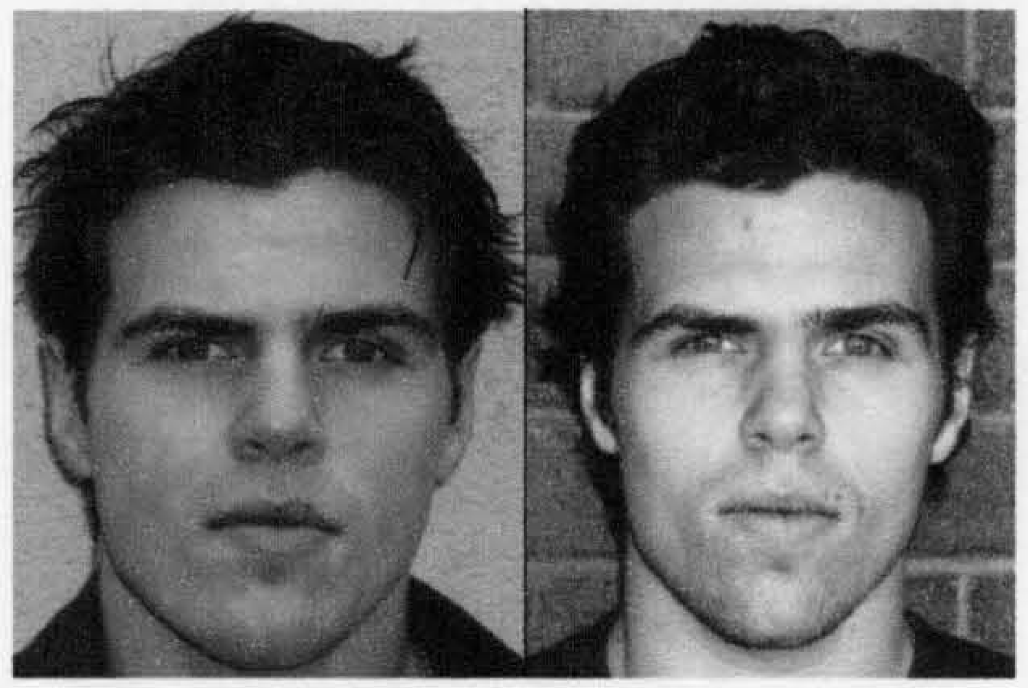
| | | | | | | |
|---|---|---|---|---|---|---|
| PH1 | PH2 | PV1 | PV2 | PV3 | PV4 | PHT1 |

***Profile angular vectors (Figure B.5)***

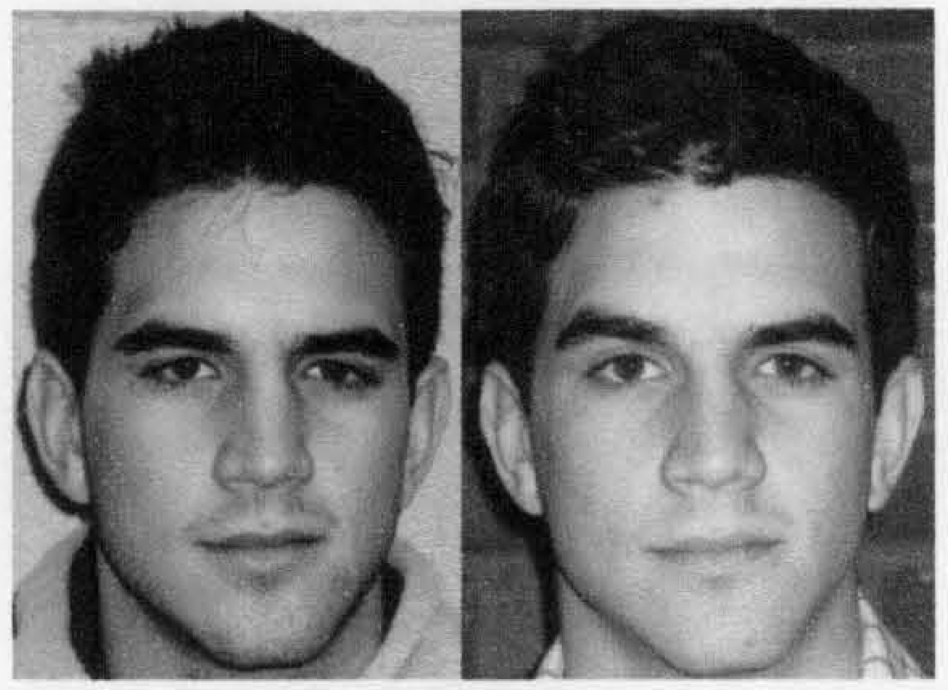
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A | B | C | D | E | I | J | K |

## Appendix D.1: Target and Probe photographs used in Chapter 10

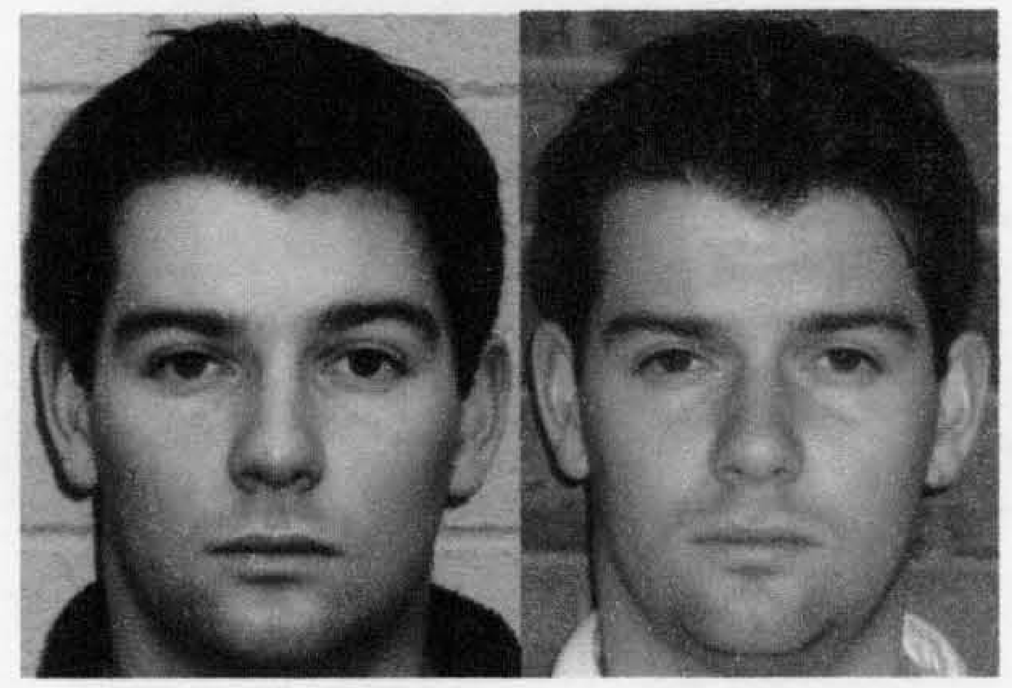
*Actor 01*



*Actor 02*

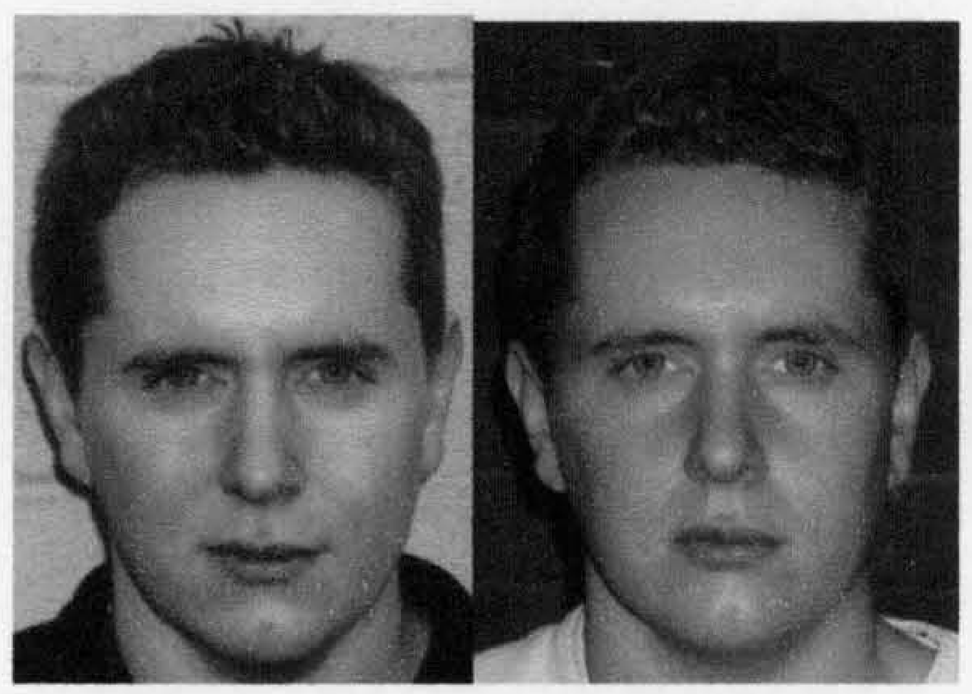


*Actor 03*



*Actor 04*

*Actor 08*



*Actor 09*

*Actor 10*

*Actor 33*

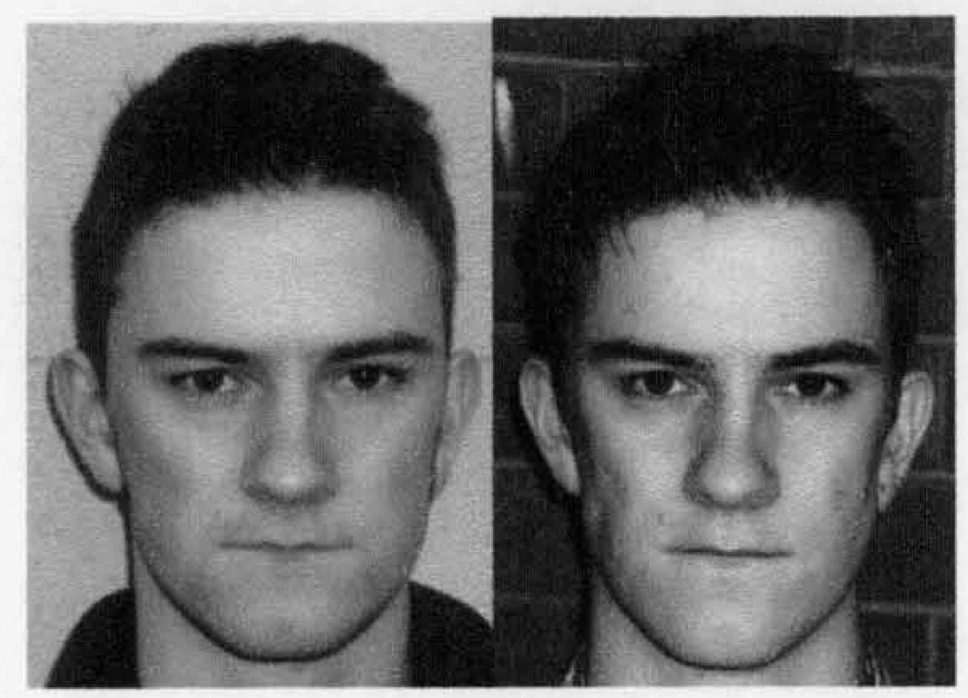
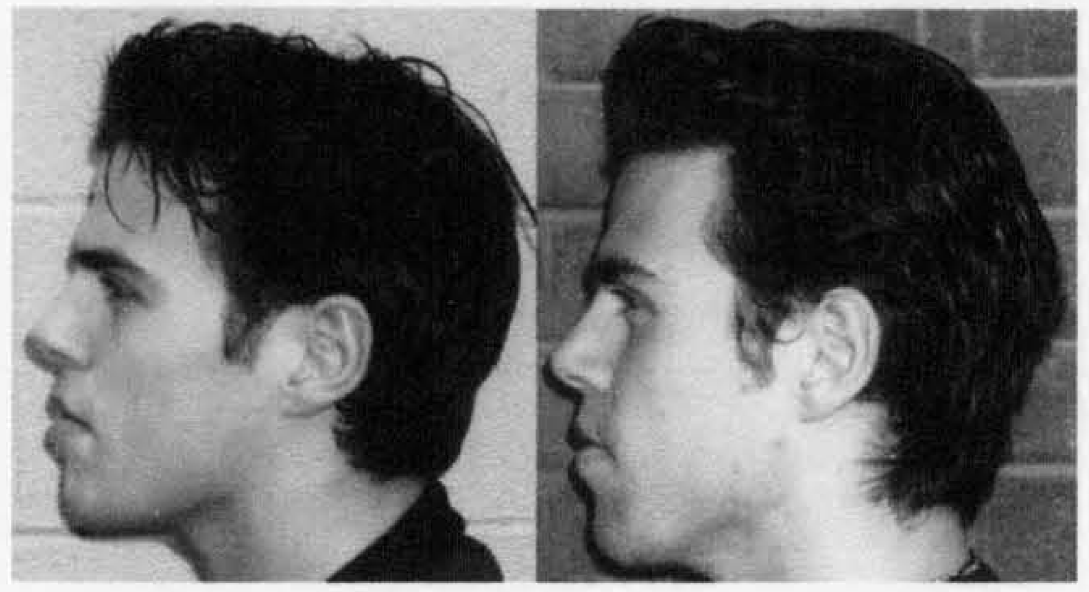


*Figure D.1: Anterior undisguised target and probe. Left image in each pair is the target the right image is the probe, taken three weeks later. See Chapter 10 for details.*

## Appendix D.2 Target and Probe profile photographs used in Chapter 10

*Actor 01*



*Actor 02*

*Actor 03*

*Actor 04*

*Actor 08*

*Actor 09*

*Actor 10*

*Actor 33*

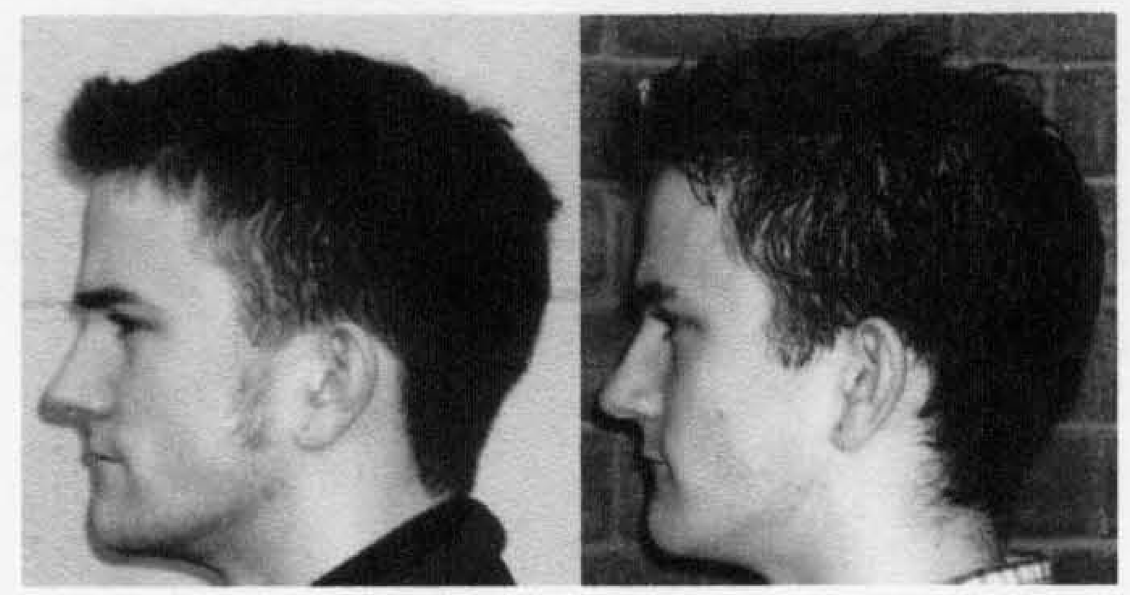
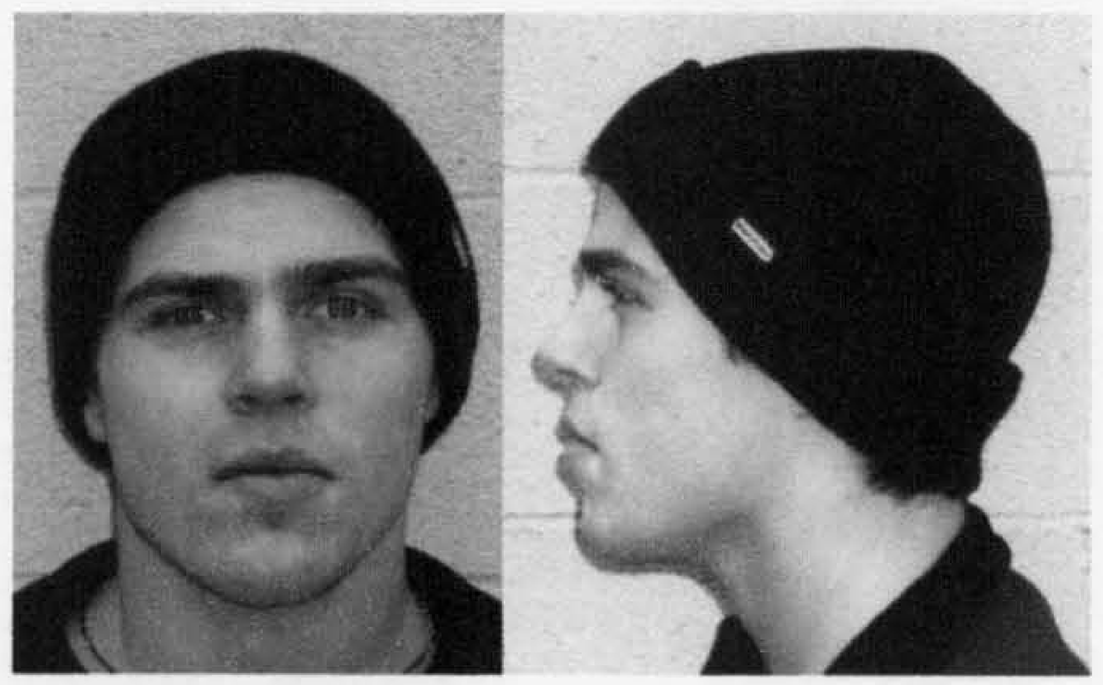


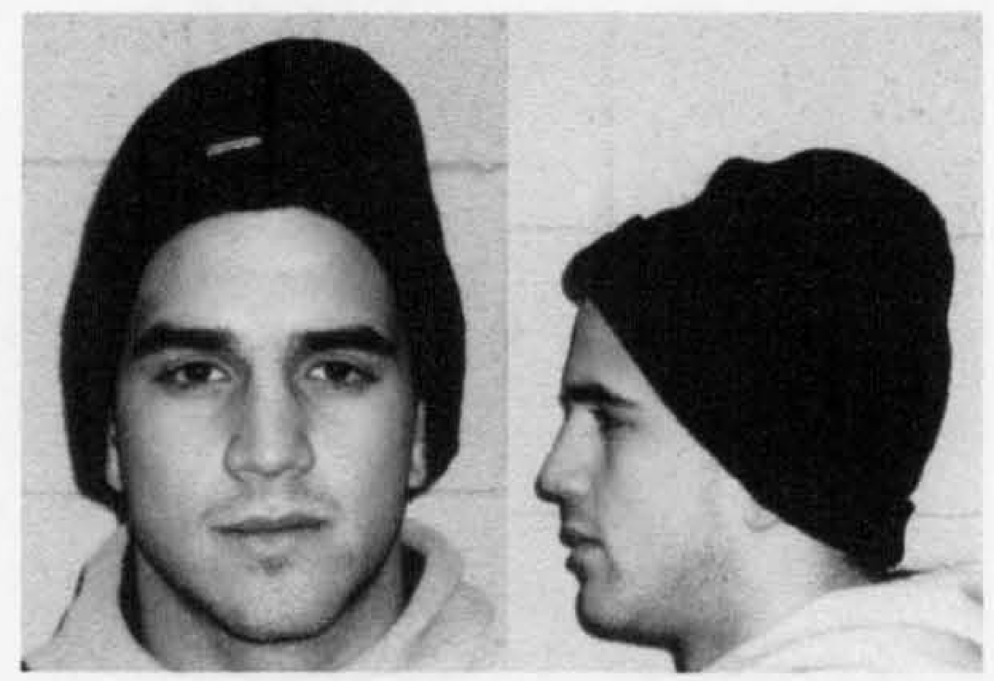
*Figure D.2: Profile undisguised target and probe. Left image in each pair is the target the right image is the probe, taken three weeks later. See Chapter 10 for details.*

## Appendix D.3: Target and Probe disguised photographs used in Chapter 10.

*Actor 01*

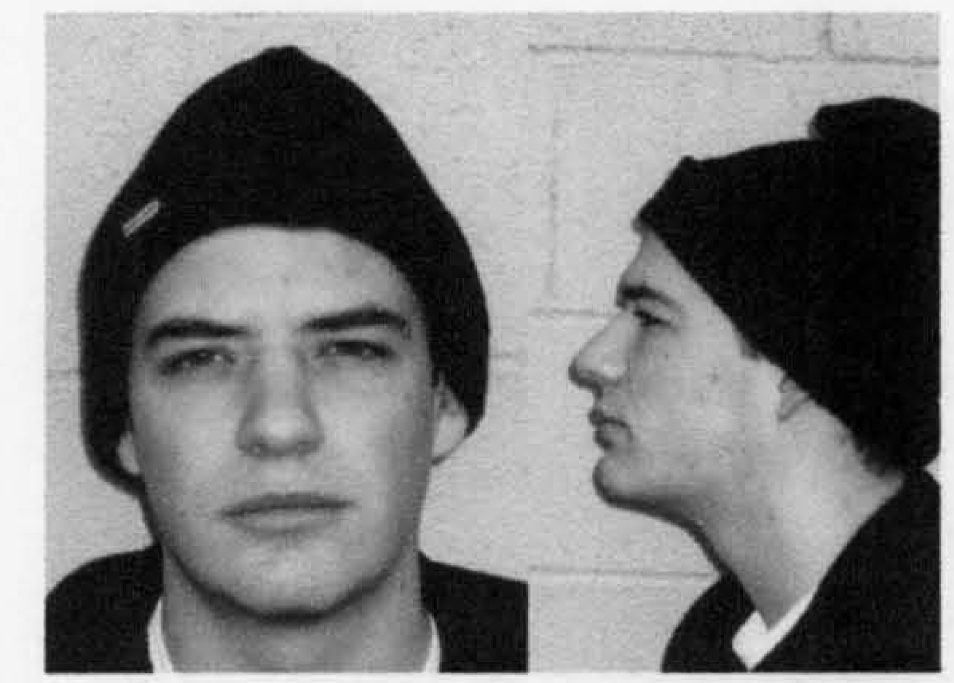


*Actor 02*



*Actor 03*

*Actor 04*



*Actor 08*

*Actor 09*

*Actor 10*

*Actor 33*

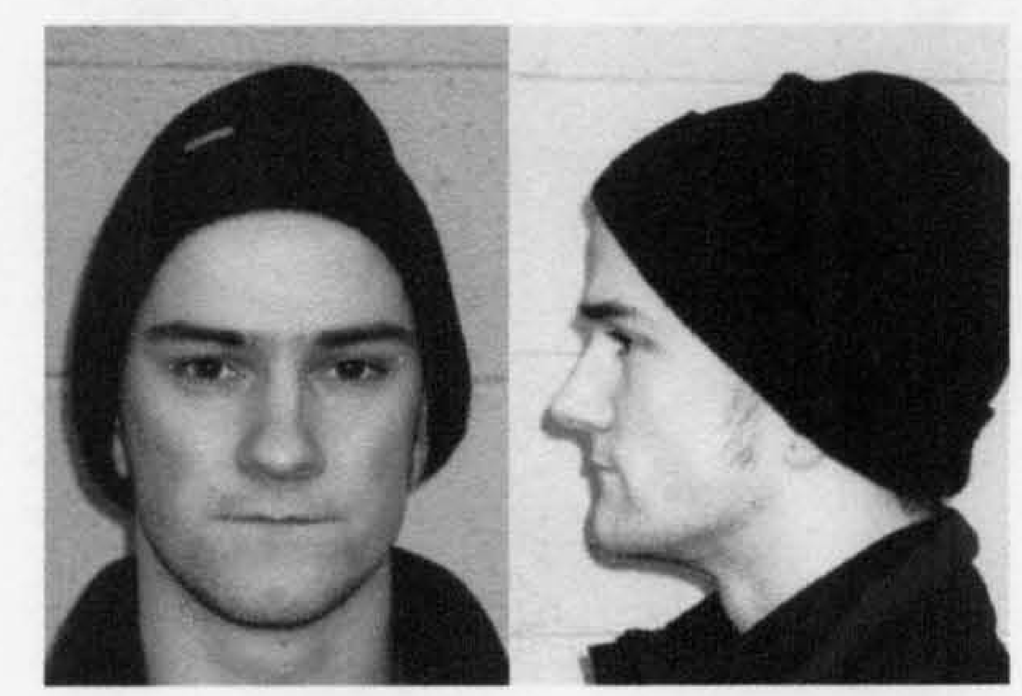


*Figure D.3: Anterior & Profile disguised probe faces: See Chapter 10 for details.*

## Appendix E.1: Anterior face database: Actors 1 to 27.

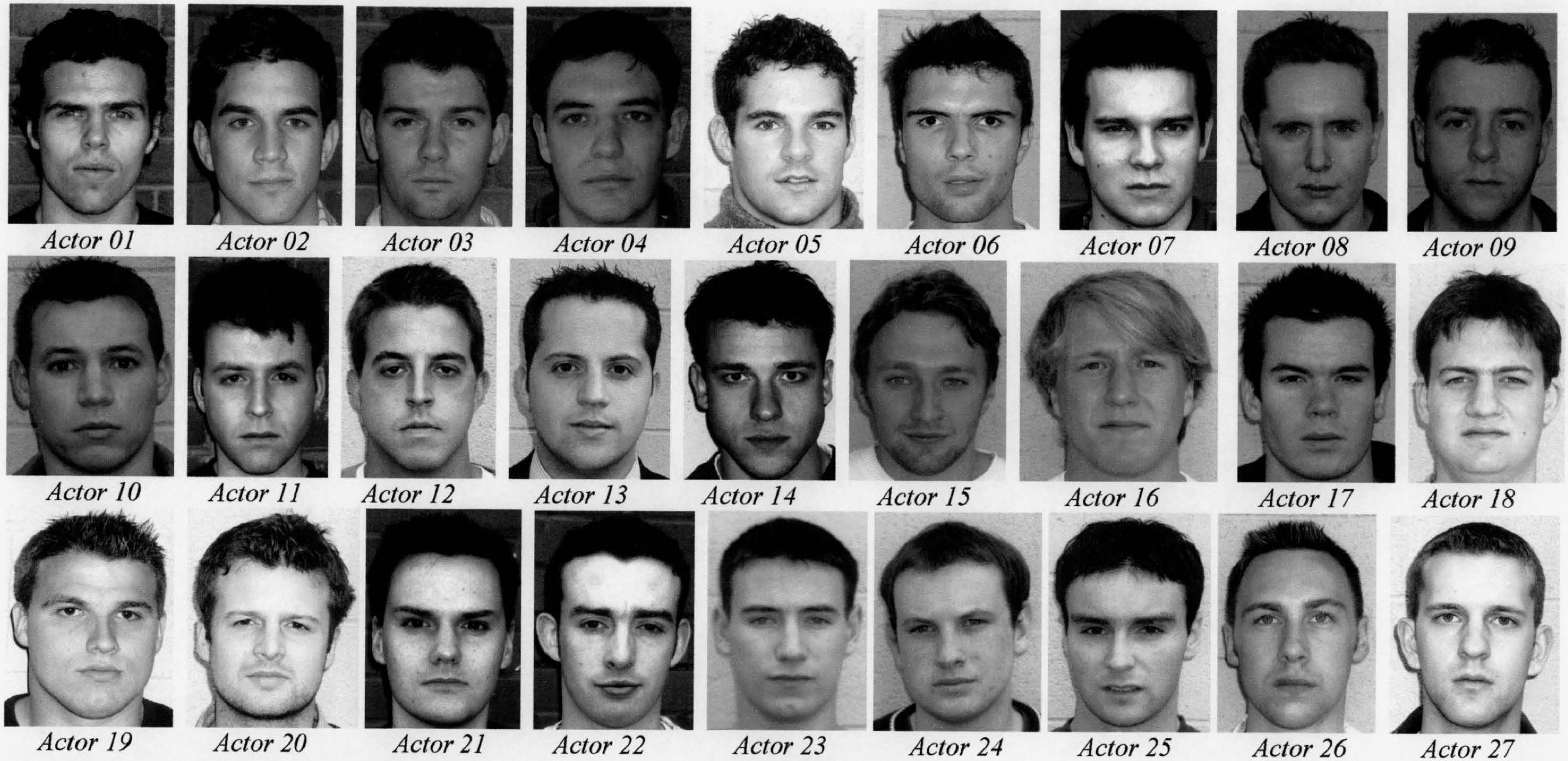



*Figure E.1: Anterior face database, Actors 01 - 27: See Section 3.3 for details.*

**Appendix E.2: Anterior face database: Actors 28 to 54.**

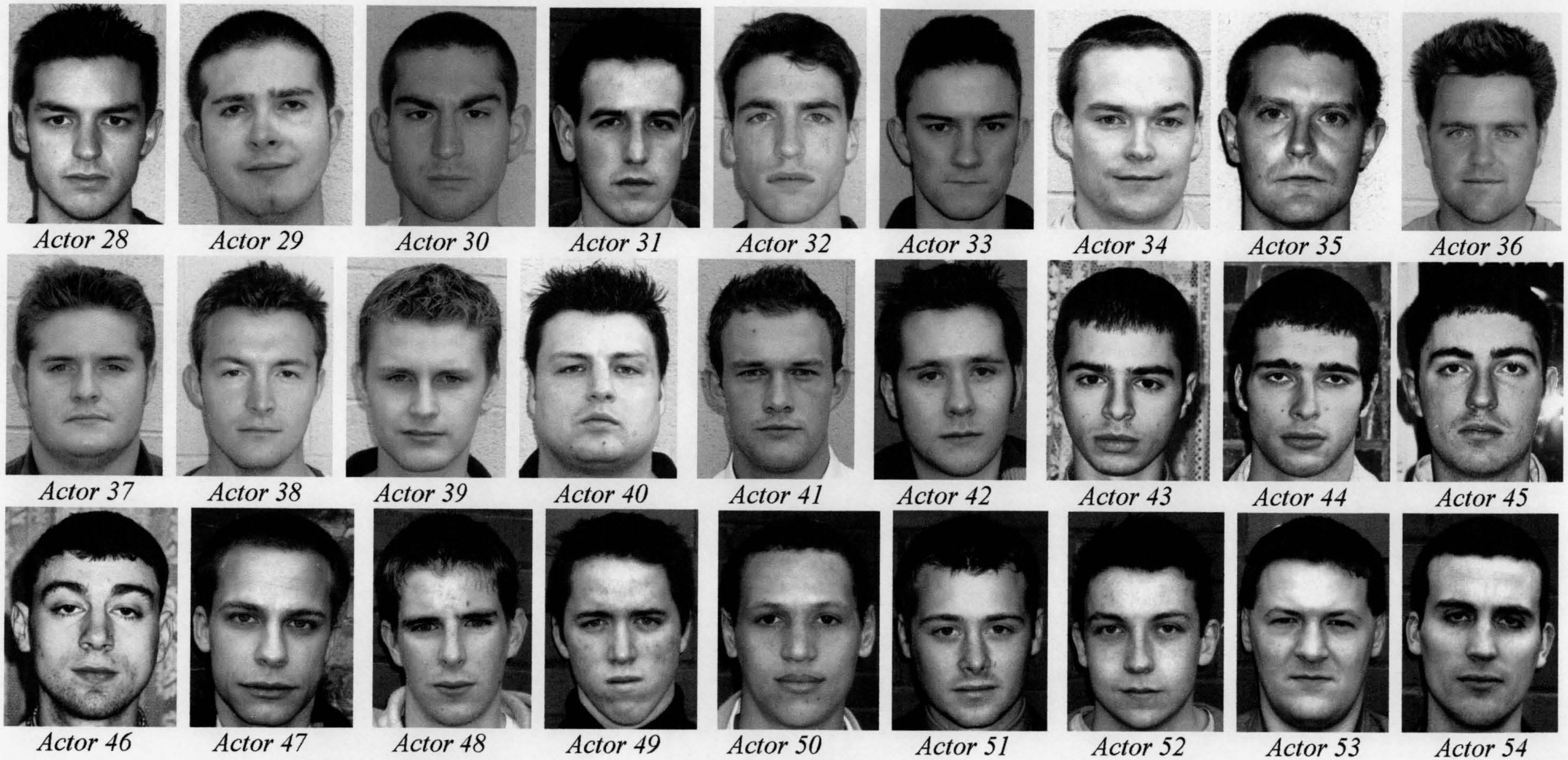



*Figure E.2: Anterior face database, Actors 28 -54: See Section 3.3 for details.*

**Appendix E.3: Anterior face database: Actors 55 to 78.**

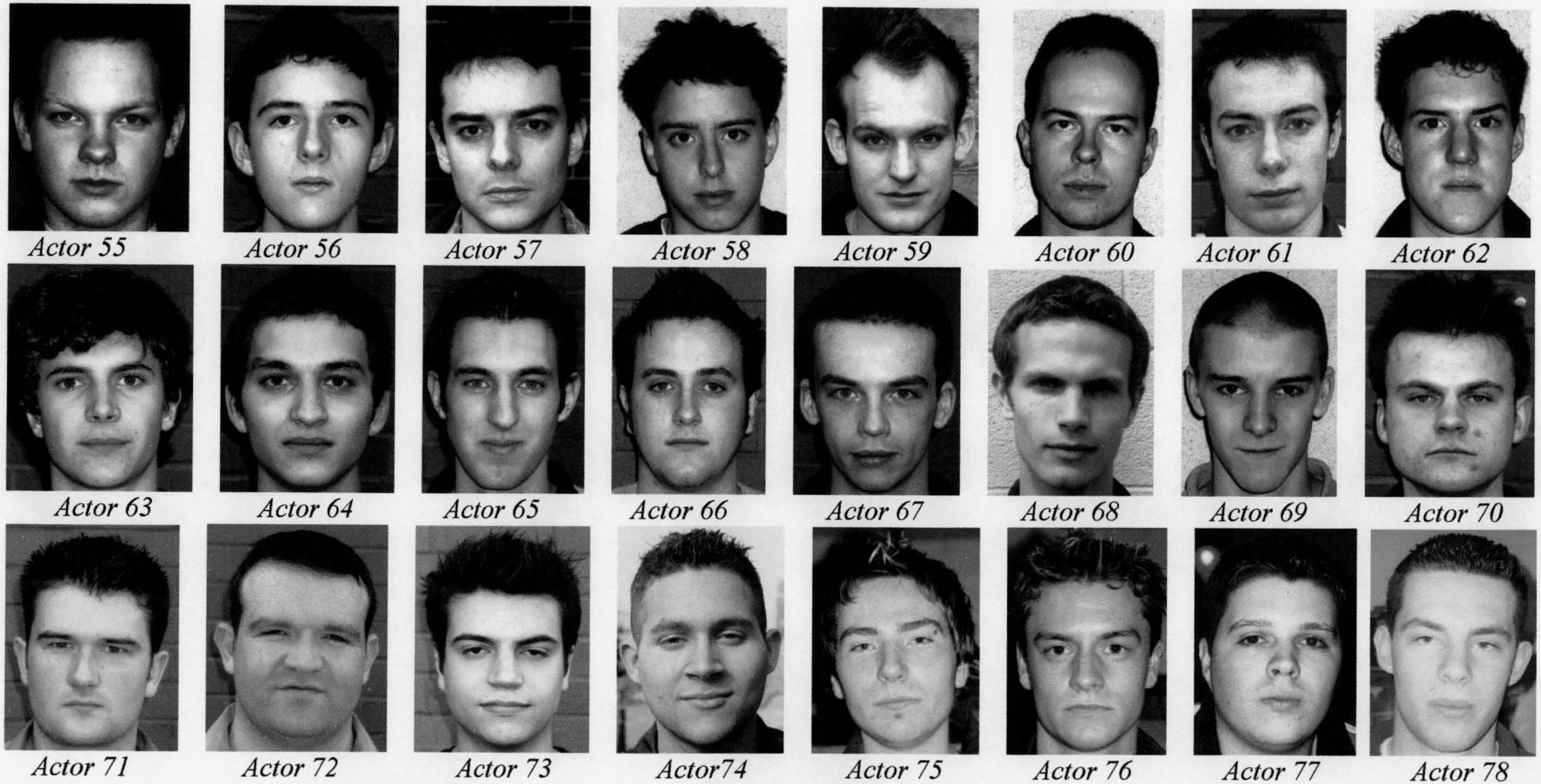



*Figure E.3: Anterior face database, Actors 55 - 78: See Section 3.3 for details.*